0541t

Community Health Cell
Library and Documentation Unit
367, "Srinivasa Nilaya"
Jakkasandra 1st Main,
1st Block, Koramangala,
BANGALORE-560 034.
Phone : 5531518

V. Mandil

Epidemiologic Methods

COMMUNITY HEALTH CELL
No. 367, Srinivasa Nilaya, Jakkasandra.
I Main, I Block, Koramangala, Bangalore - 560 034.

THIS BOOK MUST BE RETURNED BY THE DATE LAST STAMPED		

Epidemiologic Methods

Brian MacMahon
M.D., PH.D., D.P.H.

Thomas F. Pugh
M.D., M.P.H.

Johannes Ipsen
C.M., DR. MED., M.P.H.

From the
Department of Epidemiology
Harvard University, School of Public Health

DEPT. OF COMMUNITY MEDICINE
BANGALORE-34
01344
ST JOHN'S MEDICAL COLLEGE

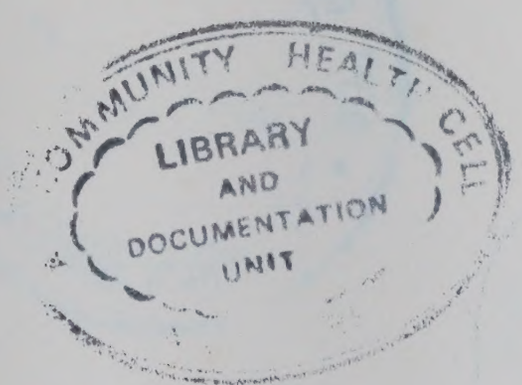
LITTLE, BROWN AND COMPANY *Boston* *Toronto*

COPYRIGHT, ©, 1960, BY BRIAN MACMAHON

ALL RIGHTS RESERVED. NO PART OF THIS BOOK MAY BE REPRODUCED IN ANY FORM WITHOUT PERMISSION IN WRITING FROM THE PUBLISHER, EXCEPT BY A REVIEWER WHO MAY QUOTE BRIEF PASSAGES IN A REVIEW TO BE PRINTED IN A MAGAZINE OR NEWSPAPER.

LIBRARY OF CONGRESS CATALOG CARD NO. 60–12344

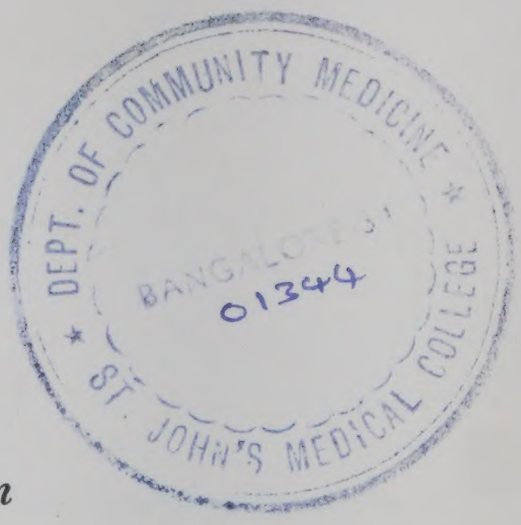
FIRST EDITION



MP100 N60
0541(

Published in Great Britain by J. & A. Churchill, Ltd., London

PRINTED IN THE UNITED STATES OF AMERICA



To

John Everett Gordon

DEPT. OF COMMUNITY MEDICINE
ST JOHN'S MEDICAL COLLEGE
BANGALORE
01344

Preface

Epidemiology is widely recognized as the basic science of preventive medicine. This recognition and the greatly increased public support of medical research have found expression in this country in a striking increase in the number of research studies and programs of an epidemiologic type. Naturally, this increase has been most marked for those diseases over which greatest concern is now being felt — arteriosclerotic heart disease, cancer and the mental disorders. While there has been repeated expression of faith in the applicability of the principles of epidemiologic investigation to noninfectious diseases, there is no doubt that the transference of epidemiologic concern from the infectious to the noninfectious diseases has necessitated appreciable modification of methods, if not of principles. We hope, therefore, that this attempt to summarize concepts that seem to us particularly appropriate to the present time will be of service.

Concern with the present by no means implies disregard for infectious disease. Outstanding problems remain unsolved in this field. In addition, our frequent use in this volume of illustrations from studies of infectious disease is testimony to the fact that the present status of epidemiology as a discipline must be credited in major part to its achievements in this area.

In spite of the increasing use of biostatistical techniques in epidemiologic studies, we have avoided, so far as possible, the duplication of material that is readily found in textbooks of biostatistics. We believe that the understanding of the principles discussed here does not require statistical knowledge. However, the reader intending to put some of these principles into practice would do well to sentence himself to a term with a book of elementary statistics, the sentence preferably to run concurrently with the present one.

Whoever wishes to investigate medicine properly should proceed thus: in the first place to consider the seasons of the year, and what effects each of them produces. Then the winds, the hot and the cold, especially such as are common to all countries, and then such as are peculiar to each locality. In the same manner, when one comes into a city to which he is a stranger, he should consider its situation, how it lies as to the winds and the rising of the sun; for its influence is not the same whether it lies to the north or the south, to the rising or to the setting sun. One should consider most attentively the waters which the inhabitants use, whether they be marshy and soft, or hard and running from elevated and rocky situations, and then if saltish and unfit for cooking; and the ground, whether it be naked and deficient in water, or wooded and well watered, and whether it lies in a hollow, confined situation, or is elevated and cold; and the mode in which the inhabitants live, and what are their pursuits, whether they are fond of drinking and eating to excess, and given to indolence, or are fond of exercise and labor.

HIPPOCRATES
On Airs, Waters and Places

Contents

	Preface	vii
1	The Place of Epidemiology in the Investigation of Disease	3
2	Epidemiologic Concepts of Cause	11
3	Ways of Classifying Ill or Disabled Persons	23
4	The Strategy of Epidemiology	32
5	Measurements of Disease Frequency	51
6	Sources of Mortality and Morbidity Data	72
7	Characteristics of Persons: Age	86
8	Characteristics of Persons: Sex and Ethnic Group	103
9	Other Characteristics of Persons	124
10	Place	141
11	Time	164
12	Birth Order and Parental Age	187
13	Analytic Epidemiology: Cohort Studies	210
14	Analytic Epidemiology: Case History Studies	233
15	Experimental Epidemiology	268
	Bibliography	281
	Index	291

Epidemiologic Methods

1

The Place of Epidemiology in the Investigation of Disease

DEFINITION

Epidemiology is the study of the distribution and determinants of disease prevalence in man.

Two main areas are indicated in the definition. These are the study of the *distribution* of disease (descriptive epidemiology) and the search for the *determinants* of the noted distribution (analytic epidemiology). The first, describing the distribution of health status in terms of age, sex, race, geography, etc., might be considered an extension of the discipline of demography to health and disease. The second, involving interpretation of the distribution in terms of possible causal factors, is the special contribution of epidemiology. Definitions of epidemiology that consider only the descriptive phase do not indicate the unique component of the discipline.

Epidemiology, like most other sciences, has developed from the study of the exotic and the unusual into an inquiry as to general principles, including explanation of the unusual. Epidemiology is now no more restricted to the study of striking outbreaks of disease than meteorology is restricted to the study of hurricanes or astronomy to eclipses of the sun. Yet a modern epidemiologist might still consider his vocation to be primarily a "medical science treating of epidemics" (Webster) in view of two recent trends in the practice of the discipline. These are an expansion

of the concept of what constitutes an epidemic, and a realization that study designed to interpret epidemic patterns cannot be restricted to periods of epidemic prevalence.

The Concept of an Epidemic

The term "epidemic" has until quite recently been almost entirely restricted to the description of acute outbreaks of infectious diseases. However, an acceptable definition of an epidemic might be "the occurrence in a community . . . of a group of illnesses of similar nature, clearly in excess of normal expectation" (American Public Health Association, 1960). This definition encompasses phenomena ranging from the classical pestilences to episodes in the behavior of many noninfectious diseases. The United States, for example, is at the present time in the grip of epidemics of at least two diseases — coronary heart disease and lung cancer — which satisfy as well as any infectious disease the criterion of being clearly in excess of normal expectation. Lung cancer is many times more common in this country than it was twenty years ago, and, although it is now the most common form of cancer in American males, there are many countries where it remains rare.

Epidemic and Nonepidemic Prevalence

A second trend has been growing realization of the value of knowledge of disease frequency and distribution during both epidemic and nonepidemic times, even when the predominant concern may be with the explanation of epidemics.

In the first place, without knowledge of nonepidemic prevalence, how can we know that an epidemic exists? How can we determine whether the frequency of a disease present in a particular population at a particular time is in excess of normal expectancy? Clearly, in order to demonstrate that a particular frequency of morbidity or mortality is in excess of the normal expectancy, it is necessary to know the frequency of the same condition in other populations and in the same population at other times.

In certain instances the existence of an epidemic is perfectly obvious. This is so when the epidemic involves a large proportion of persons, produces evidence of distinctive illness, and occurs

over a short period of time. Thus there is little difficulty in detecting epidemics of cholera, plague, smallpox or the common infections of childhood. In all these cases the disease is familiar, the difference between epidemic and nonepidemic prevalence is large, and the transition is rapid. In contrast, the current risk to an American male of dying of coronary heart disease is quite as large as the risk of death experienced during some of the major epidemics of infectious disease, but the general population remains almost unaware of the existence of an epidemic of coronary heart disease. The slow growth of the epidemic of this disease has served to conceal its enormity from the population.

Even acute epidemics may pass unnoticed if they appear in unfamiliar form. Thus, during the intense London fog of 1952 there was very limited realization of the effects of the fog on the population's health. The full effects were only appreciated when deaths for the period were counted and compared with deaths during the preceding and subsequent periods of the same year and during similar periods of previous years (Figure 1). It then became apparent that the fog was responsible for some 4000 deaths.

Second, finding of an unusually *low* frequency in a particular population at a particular time may be just as significant in understanding the causes of epidemics as the epidemics themselves. For example, the very low attack rates from cholera observed by John Snow (1855) among two groups of people in the center of an otherwise epidemic area led to a strengthening of the belief that the water supply was responsible for the epidemic. Similarly, knowledge of the virtual absence of cancer of the uterine cervix in nuns is of value in the formulation of hypotheses regarding the etiology of this disease.

Finally, in the chronic diseases, which have prolonged upswings and downswings of the epidemic wave, it may be difficult to say whether or not a given frequency qualifies as "epidemic" even if all the necessary comparative information is available. It is common to find a gradient in the relative frequencies of a disease in a number of populations. While the disease may be considered definitely epidemic in those populations at the top of the gradient by comparison with those at the bottom, the status of those in the center may be equivocal. Under such circumstances attempts to correlate quantitative statements of disease frequency

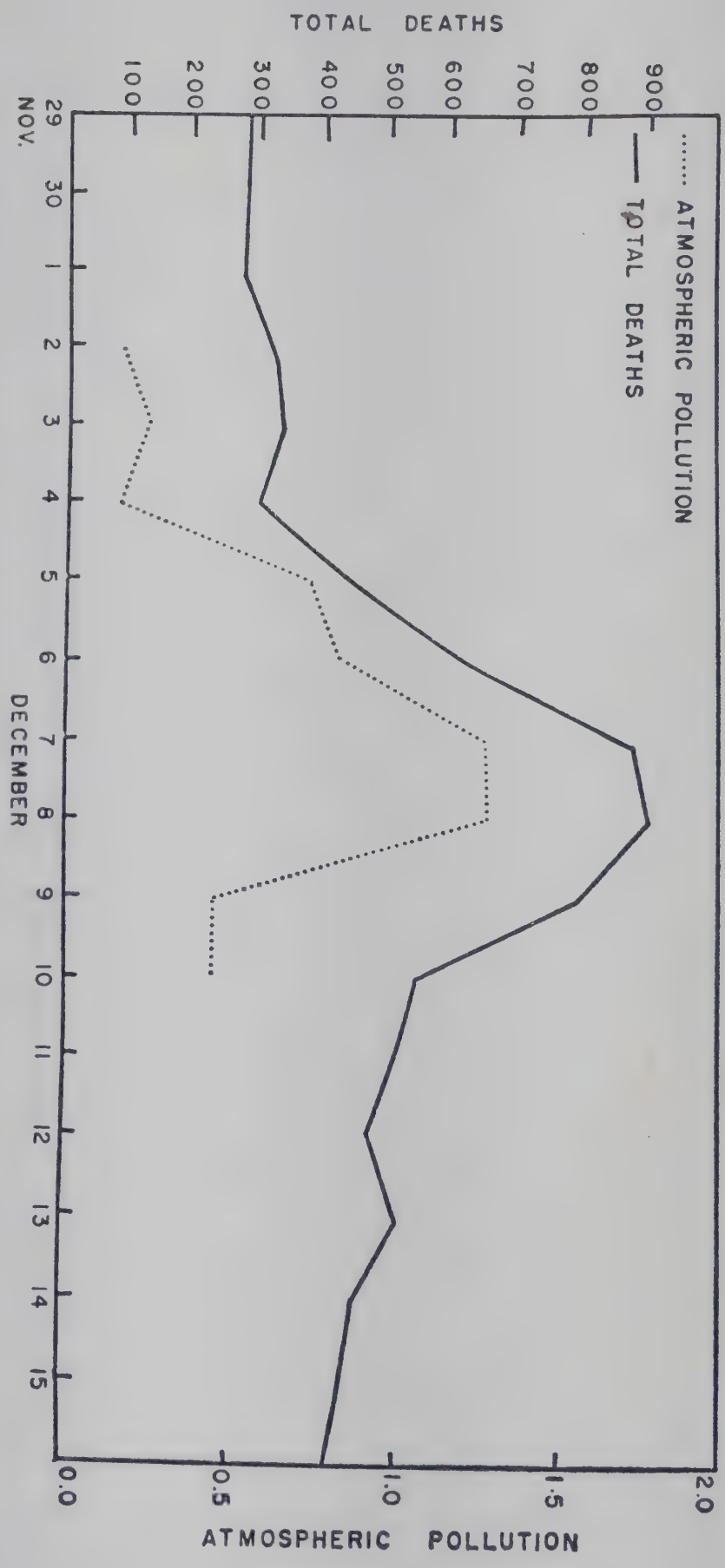


FIGURE 1. Daily numbers of deaths and degree of atmospheric pollution, London, November 29 to December 16, 1952. Atmospheric pollution is measured as parts per million of sulphur dioxide. (Data from the Ministry of Health, 1954.)

with quantitative statements of the frequency of the suspected factors are more precise than attempts to correlate the crude dichotomy of epidemicity and nonepidemicity with the crude dichotomy of presence or absence of specific factors. The former procedure has been followed, for example, in relating characteristics of the diet to rates of coronary heart disease in different populations.

These considerations have led to a great many disease frequency studies which, while having as an ultimate purpose the explanation of unusual prevalences of disease, are not oriented directly toward the study of situations known to be epidemic.

PURPOSES

Knowledge of the distribution of disease may be utilized in one of three general areas — the elucidation of causal mechanisms, the explanation of local characteristics of disease occurrence, and administrative guidance in the provision of health services.

Understanding the Causation of Disease

The most challenging purpose of epidemiology is that of identifying those components of causal mechanisms that enable the formulation of effective preventive measures. This aim encompasses a number of subsidiary, more specific purposes:

(1) Formulation, selection, or rejection of hypotheses that explain disease distribution in terms of specific characteristics or experiences of affected persons.

(2) Testing hypotheses of disease origin through special surveys or other observational studies.

(3) Testing the validity of the rationale on which control programs are based, by the use of epidemiologic data collected in conjunction with the programs. A control program may be considered as an experimental design, the subject of the experiment being the knowledge of causal mechanisms which gave rise to the program.

(4) Distinguishing "entities" of disease and disability through knowledge of disease distribution and of causal factors. On occasion, similarity of epidemiologic behavior may point to etiologic similarity between clinically distinct entities. Con-

versely, difference in epidemiologic distribution for subgroups of an apparent clinical entity may suggest that such subgroups should be regarded as separate disease entities because of probable etiologic differences.

Most frequently, epidemiologists are concerned with these aims as applied to the elucidation of the causation of disease. However, processes of human biology other than those with an end product of manifest disease have been studied. These have included prenatal and postnatal growth, multiple pregnancy, sex determination, intelligence and fertility.

It has been suggested that epidemiology should be concerned as much with those positive components of health implicit in the definition of the word as used by the World Health Organization* as with those negative characteristics that are recognized as frank disease. However, the number of widespread and serious diseases having defined clinical and pathologic characteristics but an unknown etiology appears adequate to occupy the world's epidemiologists for some years to come. Concentration of effort on these latter problems is indicated both because of the great need for knowledge leading to the prevention of these diseases and because of the practical difficulties in the statistical or epidemiologic treatment of concepts that have been neither clinically nor pathologically defined.

Understanding Local Disease Patterns

The preceding purpose dealt with the acquisition of new knowledge about the origin and nature of a disease. In contradistinction, the practicing epidemiologist is often concerned with specific local epidemics or other variations in disease frequency. Here he utilizes what is already known about the etiology of, say, typhoid fever, to explain and deal with an outbreak and to formulate preventive measures suitable to the particular community.

Of course a fine distinction between these two approaches cannot be drawn, since new knowledge may be derived during the course of routine investigation. Nevertheless, in a great deal of "practical" epidemiology, new knowledge of disease causation is not purposefully sought and its advent would be unex-

* "Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" (W.H.O., 1948).

pected. Typical of this practical epidemiology are the large numbers of investigations of localized outbreaks of food poisoning and salmonellosis undertaken by departments of health with a view to finding the local source of infection, halting the epidemic, and preventing its recurrence. At the present time, because of inadequacies of knowledge concerning other illnesses, such practical uses of epidemiology are largely confined to infectious diseases and certain industrial intoxications.

Administrative Purposes

Knowledge of the frequency of disease in a population serves a number of administrative purposes. For example, it is essential to the logical planning of facilities for medical care. Estimation of the number of hospital beds required for patients with specific diseases (tuberculosis, cancer) or for given segments of the population (infants, the aged) requires knowledge of the frequency and natural history of either those specific diseases or all diseases in those specific parts of the population. In addition, the planning of a variety of studies related to health — clinical, therapeutic, and preventive — requires, for efficiency of design, knowledge of how many cases of a particular disease are likely to be found in a given population during a given period.

Knowledge of the relative frequencies of a disease among subgroups of the population is also useful, since it enables programs and studies to be directed toward the population group manifesting the greatest concentration of the disease. For example, if facilities are limited, it assists in deciding which age, occupational, sex, geographic or ethnic group should be the target for a tuberculosis, diabetes or cancer screening program. Similar considerations apply to the choice of a subgroup of the population for any study, whatever its purpose, that requires the maximum yield of cases for a given size of population studied.

RELATIONSHIP TO OTHER DISCIPLINES

Two characteristics of epidemiology allow us to place this discipline in a logical relationship to the many other disciplines concerned with understanding the causes of disease. These are that epidemiology is primarily an applied discipline, and that its

methods are predominantly observational, rather than experimental.

Its classification as an applied discipline — that is, one concerned with the solution of immediate practical problems — implies that there are certain prerequisite disciplines, disciplines whose content is used by the epidemiologist as the starting point for his own observations. Needless to say, the content of practically any discipline may be relevant to the epidemiologic study of specific problems, but there are three disciplines that are particularly pertinent. The contributions of two, clinical medicine and pathology, to the accurate description and classification of disease form the basis for determination of disease frequency. The more clearly defined and delineated the disease or biologic process in question, the greater is the probability of successful epidemiologic observation. The third discipline, biostatistics, is necessary for determining whether observed variations in disease frequency within and between groups of the population are more than fortuitous.

The fact that the contributions of other disciplines are required in addition to those of epidemiology is implicit in the statement that the methods of epidemiology are predominantly observational. Since the most convincing test for causal relationship is usually through experiment (see Chapter 2), the methods of other disciplines are needed for more critical examination of suspected causal relationships and for the investigation of their mechanisms. Epidemiology does have a role in the experimental investigation of disease, but it is a limited one.

The disciplines whose purview it is to undertake these experimental investigations vary with the subject area within which the suspected relationship falls. The most pertinent at the present time is microbiology. However, a great deal of experimental work in biochemistry, radiobiology, and experimental pathology is aimed at the explanation of associations noted through epidemiologic observation. Since observation is in general a great deal more economical than experiment, one of the purposes of epidemiology may be considered as the identification and narrowing down of subject areas within which practical benefit is most likely to result from experimental investigation.

2

Epidemiologic Concepts of Cause

DEFINITION

All descriptive sciences take note that certain events tend to be associated with each other in time. Some of these associations between events have characteristics which lead us to speak of the relationship as cause and effect — the first event being denoted the "cause" of the second. The use of these words sometimes leads to the idea that the second event is the inevitable consequence of the first. This concept is now discredited by most schools of philosophy. The first doubts are usually ascribed to Hume (1739), who wrote: "We are never able, in a single instance, to discover any power or necessary connection, any quality which binds the effect to the cause, and renders the one an infallible consequence of the other. We only find that one does actually, in fact, follow the other."

There is nothing inevitable about causal associations, and there is no logical reason to believe that because they have been demonstrated under conditions existing in the past, they will necessarily continue in the future. Of course, the longer they have been observed and the greater their constancy in the past, the stronger will be our belief in the likelihood of their continuation in the future, but such belief rests on empirical observations rather than on any logical process of deduction.

Doubt as to the demonstrability of the basic nature of causal

relationships extends even to the observations of the physical sciences in which association between two events may be so strong as to seem inevitable within the limits of our experience.

What then are the special features of certain relationships which lead us to describe them as causal? We should note first that, the term being abstract, its meaning will not be universal. Any definition of it must take into account the purpose for which it is intended, and no definition will have equal validity in all branches of science. Epidemiology has a practical purpose in the discovery of relationships which offer possibilities of disease prevention. For this purpose, therefore, a causal association may be defined as an association between two categories of events in which a change in the frequency or quality of one is observed to follow alteration in the other. In certain instances the possibility of alteration must be presumed and a presumptive classification of an association as causal may be justified. Human genetics is a science based in its entirety on associations that can only presumptively be classified as causal.

TYPES OF ASSOCIATION

To understand the implications of the use of the term "causal association," it is necessary to describe some of the ways in which events may be related. With respect to each other, two categories of events* may be:

(A) Independent (not statistically associated)
(B) Statistically associated
    (1) Noncausally (secondarily) associated
    (2) Causally associated
        (a) Indirectly causal
        (b) Directly causal

From the above, the investigation of a relationship can be seen to progress from demonstration of statistical association to demonstration that the association is causal, and ultimately to ascertainment of its directness.

* By a "category of events" is meant all the single events having the specific characteristics necessary to warrant their classification together; for example, "multiple sclerosis" as distinct from the single event "a case of multiple sclerosis."

Statistical Association

If a category of events has a frequency x and another category has a frequency y, the two types of events will occur together with a predictable frequency, equal in fact to the product of their individual frequencies, xy. Suppose one of these categories to be a disease and the other to be an attribute or experience of man: there will be a certain number of individuals who exhibit both the experience and the disease. While the single events may be said to be associated (in the nontechnical sense) in such individuals, these instances of coexistence are not evidence of statistical association. Statistical association, or merely "association" in the scientific sense, implies that the proportion of individuals exhibiting both events is either significantly higher or significantly lower than the proportion predicted on the basis of the separate frequencies of the two categories. If the two categories of events are each measured on quantitative scales rather than in terms of frequency, a correlation significantly different from zero is accepted as evidence of association.

While the content of the last paragraph is obvious enough, its main implication is not widely accepted. Since the demonstration of statistical association requires comparison of observed with predicted coexistence of categories of events, and since statistical association is a necessary characteristic of cause, it is apparent that cause can only be determined in statistical terms. In other words, it is never possible to determine whether two single events are causally related; such determination can only be made with respect to categories of events.

For example, suppose that 100 persons are inoculated with a vaccine against a certain infectious disease and that 100 persons undistinguishable from the first group receive a placebo. During a subsequent epidemic, although the two groups have similar exposure, 20 of the vaccinated persons and 50 of the unvaccinated contract the disease. Since this difference is unlikely to be due to chance, we would say that a statistical association exists between vaccination and remaining free of the disease and, in view of criteria discussed below (page 15), that this association is probably a causal one. However, it is not possible to say that vaccination

caused any individual person in the vaccinated series to remain disease-free, since there were instances of vaccinated persons who contracted the disease and of unvaccinated who remained free. Within the framework of these results, it is even possible that there were persons who contracted the disease *because* of vaccination, even though the over-all tendency was in the opposite direction.

Of course, information from a group experience may suggest the probability of the existence of causal association in an individual instance. The stronger the association between the two categories of events revealed by the group experience, the more likely is the assumption of causal association in a specific instance to be correct. Thus, if the disease frequency in the unvaccinated series had been 99 per cent and that in the vaccinated series 1 per cent, there would be a very high probability that the absence of disease in any one vaccinated individual was related to the vaccination. The statement that the vaccination of any one individual was causally related to his freedom from disease would, in such circumstances, be likely to be correct. The validity of the statement depends, however, on the total experience, and not on any observations made on this one individual (other than that he was one of those who were vaccinated and remained free of disease).

Causal and Secondary Association

Only a minority of statistical associations are causal within the sense of the definition, which requires that change in one party to the association alters the other. The large number of statistical associations which do not satisfy this requirement may, for want of a better term, be described as secondary. Secondary associations usually result from association of both categories of events with a third category. For example, if category C is causally associated with both category A and category B (i.e., C precedes and influences both A and B), A and B will necessarily be associated statistically. However, the association between A and B is noncausal, since there is no prospect of altering B by manipulating A. This type of association is common. For example, injection of neoarsphenamine (A) in outpatient clinics for venereal disease has been noted to be associated with jaundice — "salvarsan icterus"

(B). For a long time, as the name indicates, the drug was regarded as the cause of the icterus, until it was discovered that the association was the result of causal association of both icterus and injection of neoarsphenamine with a third factor — treatment for syphilis (C).

Once a statistical association has been demonstrated, how can it be determined whether or not it is causal — if A is altered, is B affected? The most satisfactory procedure is by direct experiment. However, this procedure, while freely available to the laboratory worker, is infrequently available to the epidemiologist. Quite apart from the general limitations on human experimentation, experiments in disease prevention are difficult to perform, frequently involve considerable disturbance of personal and social behavior, and are expensive. They are unlikely to be undertaken unless the probability of a favorable outcome is very high. In certain associations, the presumed cause may not be susceptible to manipulation and therefore to experiment; while knowledge of whether the relationship is or is not a causal one may in this event be of little immediate value, it may nevertheless be important from the standpoint of the development of knowledge.

In the absence of experimental proof, three types of consideration are useful in distinguishing between causal and secondary associations:

(1) Time sequence. Associations in which the related events occur simultaneously or with varying precedence cannot be considered causal. In a causal relationship it is necessary that the sequence of individual events be consistent. In situations in which the sequence of events cannot be determined precisely (a not uncommon situation in chronic disease), at least the possibility of consistent sequence must exist for a relationship to be considered causal.

(2) Strength of the association. The stronger the association between two categories of events (for example, the higher the ratio of the incidence of B following A to the incidence of B without A), the more likely the association is to be causal.

(3) Consistency with existing knowledge. Clinical, pathologic and other knowledge of the disease in question and knowledge of other pathologic or physiologic effects of the supposed causal factor, leading to a judgment as to whether or not a causal

hypothesis is "reasonable," play an important part in the acceptance of such a hypothesis. On the other hand, collateral evidence may suggest a satisfactory explanation of the association other than as the result of a causal relationship.

The assessment of a relationship as causal in the absence of direct experiment is neither easy nor objective. Differences of opinion resulting from subjective interpretation and assembly of evidence are common. Caution in presuming relationships to be causal is laudable. On occasion, however, such caution appears to be carried to an unrealistic extreme, the implications of sound evidence pointing to the causal nature of an established relationship being stubbornly resisted. In situations in which the derivation of experimental proof is either impracticable or unethical, complete evidence will never be at hand, and it may be imprudent to await it.

Direct and Indirect Causal Association

At the beginning of the previous section (page 14), we discussed a statistical association between two categories of events that was explicable in terms of association of both with a third category. This type of secondary association must be distinguished from causal associations in which a third variable occupies an intermediate stage. Thus, if A is causally related to D (A being the cause and D the effect) and D is causally related to B (D the cause, B the effect), there will be a causal relationship between A and B, although the association is said to be indirect. For example, treatment for syphilis is not of itself productive of icterus, but it is one of the factors associated with use of unclean syringes. Further investigation of "salvarsan icterus" indicated that the unclean syringe component rather than the treatment of the syphilis was responsible for the icterus. However, since a certain number of cases of icterus would presumably be prevented by failure to treat syphilis, the association of syphilis treatment with icterus is causal, although indirectly so.

The distinction between direct and indirect causal relationships is a relative one. Actually, all causal relationships are indirect, apparent directness depending on the limitations of current knowledge. In the example above, the association of icterus with syphilis treatment was indirect, and that with the use of

unclean syringes direct. However, further investigation revealed that the icterus was associated not with unclean syringes *per se*, but with the injection of small remnants of human serum associated with the use of unclean syringes. This discovery resulted in a change of the name of the condition to "serum hepatitis." However, the icterus was further found to be associated directly, not with serum, but with the presence in the serum, on occasion, of hepatitis virus. Thus the association with the virus is currently the direct one and that with serum an indirect one. Further studies might in due course reveal what specific attributes or molecular components of the virus might be considered more direct causal factors. Knowledge of causal mechanisms will never be refined to the degree that makes it possible to state that *this* is the ultimate direct association and that no other, more direct associations intervene.

A goal of complete knowledge would require study of causal associations until the most direct causal mechanisms that can be observed are identified (i.e., ad infinitum). However, the practical significance of causal associations in the development of preventive programs does not necessarily depend on the degree of directness. First, more direct associations may not yet have been identified and there may therefore be no choice but to make use of obviously indirect associations. For example, knowledge of the association of freedom from scurvy with diets containing fresh fruits and vegetables was put to practical use hundreds of years before the isolation of vitamin C, and prevention of smallpox certainly antedates modern virology by several centuries. A similar situation exists with respect to the association of cigarette smoking with lung cancer at the present time. Second, intermediate direct causes may not be susceptible of economical alteration, even if known, whereas the indirect ones are. For example, preventive measures against serum hepatitis are directed against poor syringe hygiene and not specifically toward removal of the hepatitis virus. Furthermore, even seventy years after the discovery of the microorganisms associated with enteric disease, preventive measures are directed, at least in the United States and Europe, primarily toward the provision of clean water and food.

THE WEB OF CAUSATION

We have discussed the types of association that may exist between two categories of events. In fact, effects are never dependent on single causes. The concept of "chains of causation" although common has the defect of oversimplification. In Figure 2 are shown some of the components that enter into the causal association between treatment for syphilis and serum hepatitis. When it is considered that only a few of the major components are shown, that these are indicated as broad classes of events rather than as the multiple minor events comprising each class, that each component shown is itself the result of a complex genealogy of antecedents, and that the myriad of effects of these components other than those contributing to the development of icterus are not shown, then the "chains" of causation become difficult to elucidate and the whole genealogy might be thought of more appropriately as a web, which in its complexity and origins lies quite beyond our understanding.

Fortunately, to effect preventive measures, it is not necessary to understand causal mechanisms in their entirety. Even knowledge of one small component may allow some degree of prevention.

The concept of the "chain" mechanism is that many variables may be related to a single effect through a "direct-indirect" mechanism in which C is causally related to D, D to E, E to F, and so on until finally maybe Q plays an important part in the development of the disease. We may note that many variables other than those directly in the chain enter the genealogy at each step of the chain. Consequently, the longer the chain the weaker the association. A preventive attack on the disease depends on finding an element in the chain which can be eliminated and which is sufficiently close to Q in the mechanism for its elimination to have a substantial effect on Q. Thus the injection of hepatitis virus in venereal disease clinics can be prevented at a number of different points in the chain, and wherever the chain is broken the disease will be prevented, even though no attempt has been made to alter, for example, the immunity status of the patients injected.

The effect that the elimination of Q will have on the total incidence of the disease will depend on the importance of Q relative

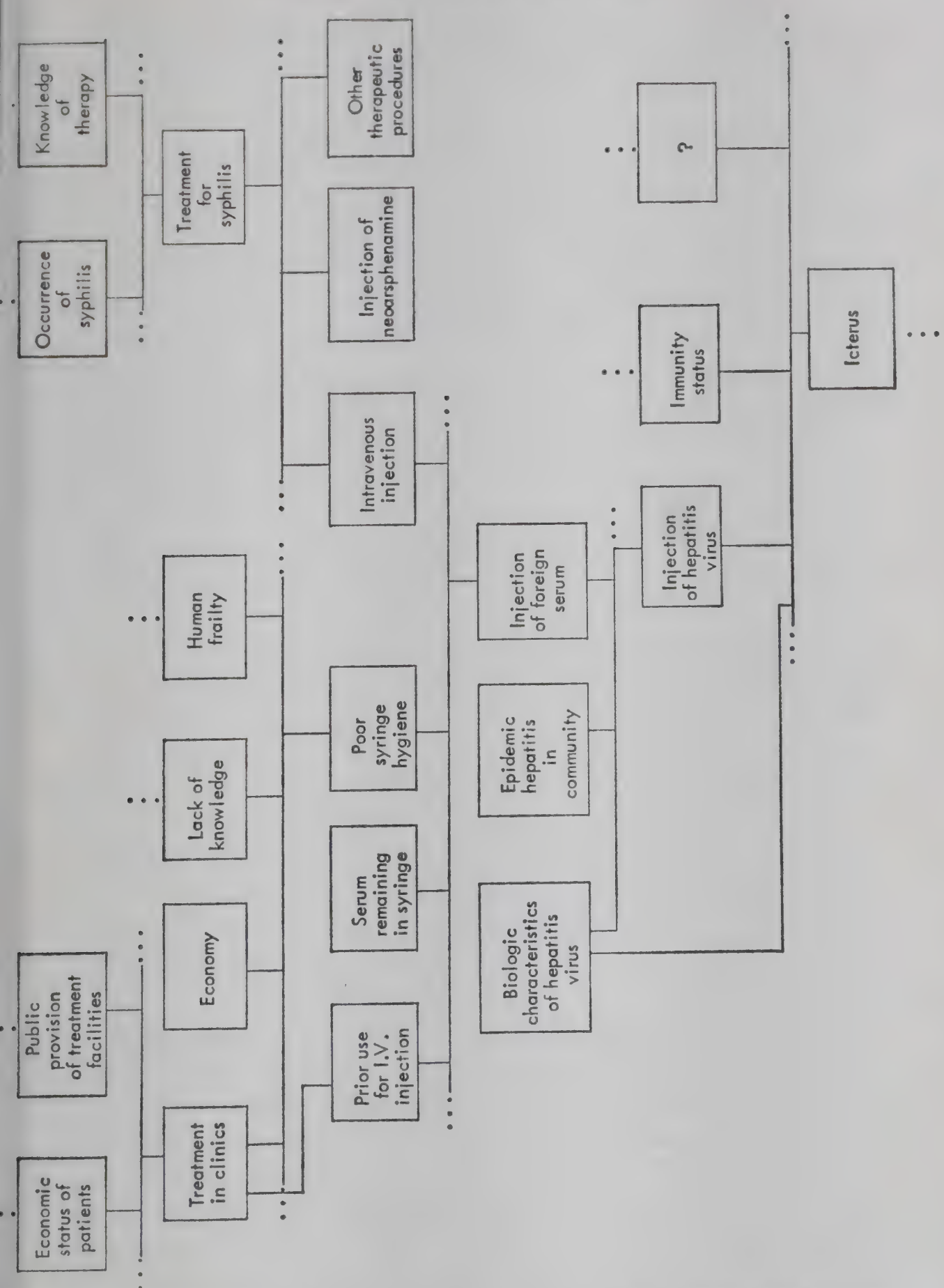


FIGURE 2. Some components of the association between treatment for syphilis and jaundice.

to other factors that, descended from independent sources, also are causally related to the disease. The effect of these independent variables may be thought of as additive: the elimination of any one of them may have an appreciable effect on the disease frequency, but nevertheless the effect of the other factors remains. For example, the elimination of serum hepatitis associated with treatment for syphilis will have only very indirect effects on the disease arising from other sources such as blood transfusion.

Ideally, one would search for components of the causal mechanism that are *necessary* causes — causes without which the disease does not occur — even though such causes may not be of themselves sufficient to produce the disease. Thus, *Mycobacterium tuberculosis* may be thought of as the necessary cause of tuberculosis, and other factors such as age, nutritional status, sex, genetic constitution, and so on, as additive "contributory" factors. However, this distinction between causes and contributory factors is arbitrary and often quite meaningless. Most commonly, the chosen factor is the necessary cause only by definition, in the sense that automobiles are the necessary cause of automobile accidents. Thus *M. tuberculosis* may be the necessary cause of tuberculosis as it is presently defined, but had medical knowledge developed differently, this pathologic state might have been included in the definition of a specific nutritional defect. In that case the nutritional defect would have been the necessary cause and the bacillus a contributory factor (while there is a definite statistical association between the tubercle bacillus and tuberculosis, the bacillus is frequently found in individuals who are not ill, and sometimes is not found in individuals with an illness clinically indistinguishable from tuberculosis).

The method of quantitative estimation of the disease incidence attributable to identified causal factors, as described in Chapter 13, seems preferable to semantic exercises aimed at a hierarchic classification of causes. Once again we may note that the practical implications of a causal association may be influenced by considerations other than quantitative estimates of the "importance" of the association. Thus the role of microbiologic agents in infectious disease has been considered predominant over that of genetic factors, perhaps in part because their importance became evident earlier but primarily because it is a good deal easier

to control human water supply and to eradicate insect vectors than to breed genetically resistant populations.

CAUSATION AND PREVENTION

The etiology of a disease may be thought of in two parts: (1) those events occurring prior to some initial bodily change (inception of the disease), and (2) those mechanisms operative within the body which lead from the initial change to the characteristic manifestation of the disease. Thus, one can conceive of a "basic" association or group of associations of high strength and high specificity which involves the effect of a specific environment on a specific type of individual. For example, it may be that penetration of a specifically predisposed marrow cell by ionizing rays will lead to genetic mutation in a very high proportion of cases, and, if circumstances are appropriate for the development of the cell, to the patient's becoming leukemic. The reaction of the marrow cell to the radioactivity may be considered the direct causal association (although it is no more direct than our methods of examination are capable of detecting), and we may describe: (1) the series of prior events which brought the radioactive particles in contact with a susceptible marrow cell, and (2) the series of events initiated by the mutation which finally resulted in the death of the patient.

This is, of course, a gross oversimplification of the development of a disease. To take a historical analogy, we speak of an "epoch" as an event which resulted in a significant change in the course of a country, for example a war or a revolution. Actually, such events consist of an infinite number of different events interacting with each other but each having its own distinct antecedents. Interference with any one of these individual events may critically influence the course of the whole epoch. Nevertheless, viewed from a temporal or geographic distance, an epoch becomes thought of as a single event either because of the blurring associated with the long-range view or because of the arbitrary selection of some point as the "turning point" of the epoch. We can then speak of the periods preceding and succeeding the epoch, as, for example, the prewar era and the postwar era, and describe the events leading up to the war and the results following it.

Prevention (and epidemiology) are concerned predominantly with those successions of events which result in the exposure of specific types of individuals to specific types of environments leading to or aggravating bodily changes in those individuals (preinception era). Therapy is concerned with the mechanism initiated in the basic bodily changes which finally results in the manifest signs and symptoms of the disease (postinception era). This distinction may lead to differences between practitioners of prevention and of therapy in their concepts of what constitutes an important causal association. For example, the therapist may be satisfied to know that deficiency of pancreatic secretion is an important cause of diabetes mellitus, since if this deficiency is remedied the symptoms of the disease are significantly modified. Such knowledge is inadequate for prevention, however, for which purpose more pertinent questions would be what caused the pancreatic deficiency.

If knowledge of causal associations is sought predominantly for its practical application, there is one other characteristic of such associations that it is desirable to determine: the side effects of alteration of the cause. For, just as any effect has multiple causes, the alteration of any cause may be expected to have multiple effects besides the one intended, and in any prevention program the side effects of alteration of the cause must be acceptable. On the one hand, a certain loss of character in the taste of public water supplies is apparently an acceptable price to pay for a typhoid-free existence, but on the other, the citizens of many communities have expressed their view that a predictable reduction in the incidence of dental decay is not worth the conjectured risks that fluoridation of water supplies might entail.

Knowledge of causal associations that do not offer preventive possibilities, either because the cause is unalterable or because the side effects are unacceptable, is nevertheless important. Knowledge of such associations is necessary for general understanding of the nature of disease, and, in addition, allows more accurate studies directed toward the identification of causal associations that do offer preventive possibilities. Thus, association with age can hardly be thought of as an association with preventive application, but if epidemiologic studies of chronic disease did not take this association into account they would be meaningless.

3

Ways of Classifying Ill or Disabled Persons

Classification of disease is sometimes a prerequisite and sometimes a goal of epidemiologic study. To understand the methods of epidemiology it is necessary to have some concepts of the methods and purposes of the classificatory process.

Classifications are arrangements of things into groups so that the members of one group can be distinguished from those of another. The things arranged may be tangible objects (e.g., books, jewels, or persons) or abstract concepts (e.g., countries of birth, religious experiences, or even classifications) (Hempel, 1959).

The purpose of classification is simple: while it is not possible to form generalizations from an individual thing, consideration of similar things in a group does permit generalizations. Study of things in a group may be productive of generalizations that are (1) explanatory of the origin of the group, (2) predictive of changes in the group with the passage of time subsequent to its delineation, or (3) both explanatory and predictive.

Classifications themselves are traditionally designated as artificial or natural. Linnaeus's system of dividing plants into twenty-four classes according to the stamens is often cited as an example of an artificial system, while those classifications which fit in with current evidence on plant evolution are said to be natural.

The distinguishing of classifications into these two types does not seem particularly realistic. The level of existing knowledge largely determines what criteria are used to arrange things into groups. In a time not yet influenced by Darwinian theory it seems not at all unnatural that differences between plants in the characteristics of the stamens were considered important. Further, the particular interest of the observer determines the kind of criteria selected to identify groups; for example, a landscape gardener may find it useful for his purposes to classify bushes as large and small.

Groupings of Ill or Disabled Persons

The arrangement of ill or disabled persons into groups so that the members of one group may be distinguished from the members of another (with respect to specified criteria connected with the illness of each individual) is essential to the concept of "separate" diseases or disabilities. A group of people so classified is said to suffer from *a* disease or *a* disability. Study of such a group may lead to generalizations of therapeutic, epidemiologic and preventive importance.

The criteria by which the members of one group are distinguished from those of another depend, of course, on available knowledge and techniques. At the moment two rather distinct types of criteria are used to delimit one disease or disability from another:

(1) Manifestational criteria. The ill persons are grouped according to their similarity with respect to several simultaneously present manifestations of illness such as symptoms, signs, changes in body fluids or tissues, physiologic function, thinking, behavior, concurrence in time, place or person, or prognosis. Examples are fracture of the neck of the femur, diabetes mellitus, idiopathic mental retardation, the common cold, Devonshire colic, miner's asthma, and cervical cancer.

(2) Experiential criteria. Here the grouping of the ill persons depends on similarity with respect to a specified type of experience begun at a time preceding the illness, and often said, loosely, to be *the* cause of a disease. Examples of experiential criteria in current use are difficult birth, combat, inhalation of silica dust, and internalization of the tubercle bacillus.

Theoretically it is possible that the same group of ill individuals might be identified through the separate use of either type of criterion. For example, it is conceivable that the manifestations of measles (rubeola) are so obvious and distinctive that the group of persons identified on the basis of similarity of symptoms and signs might contain just those persons identified separately on the basis of encountering the measles virus for the first time.

In most instances, however, such a correspondence following separate application of the two types of criteria would be unexpected. Figure 3, representing an array of ill persons, is a simple

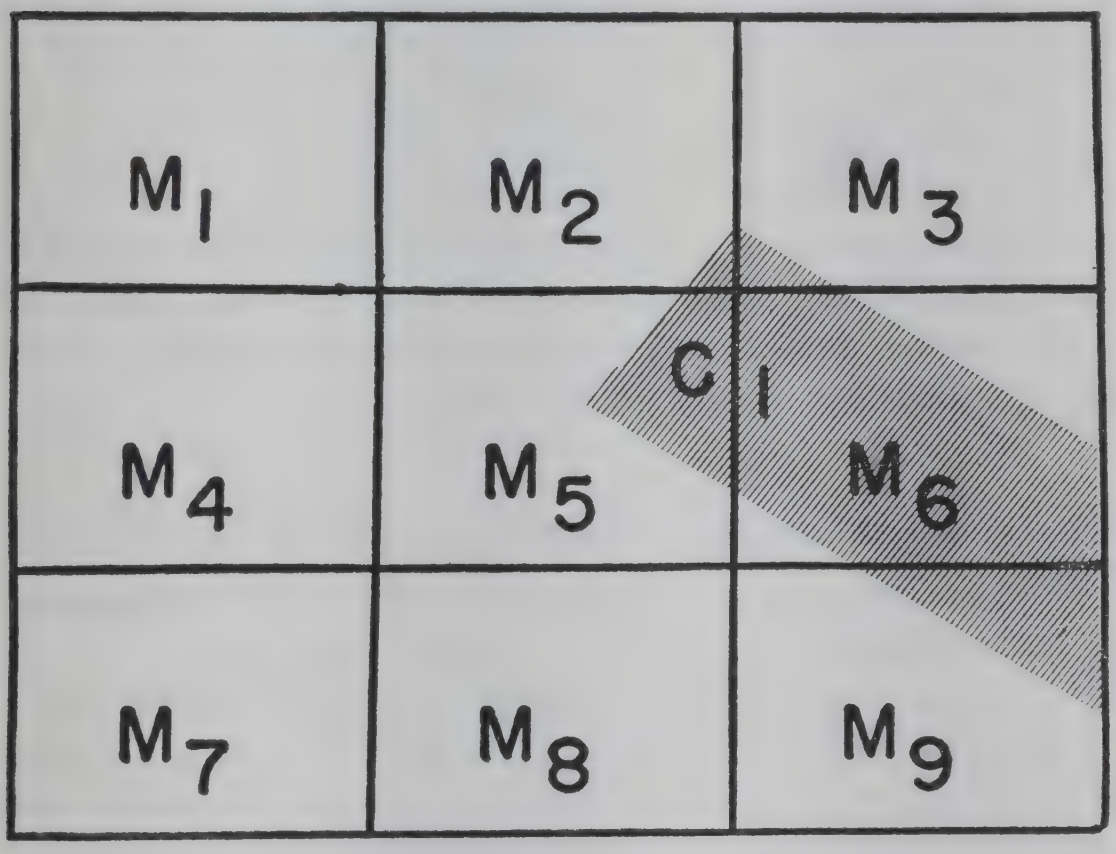


FIGURE 3. Model for classification of ill persons according to manifestational (unshaded) and experiential (shaded) criteria.

model intended to illustrate this concept. The square areas M_1, M_2 . . . M_9 are groups of ill persons, the members of one group having been distinguished from those of another only by manifestational criteria. The shaded area, C_1, represents ill persons identified separately on the basis of a specified experience. It may be noted that the correspondence is not perfect: ill persons included by the experiential criterion although mainly belonging

to M_6 are also derived from four other manifestational groups, M_2, M_3, M_5 and M_9.

This lack of complete correspondence seems to result from two facts: (1) a single specified experience (selected from the entire causal web) usually has polymorphous effects; and (2) a particular manifestation at times arises from diverse experience. For example, some pedestrians injured by an automobile may present a fractured tibia, others a simple laceration of the skin, and so on; but among all persons with a fractured tibia only a fraction have experienced a blow from an automobile.

The groups identified by either type of criterion may properly be regarded as entities of disease or disability. It should be borne in mind, however, that in both cases the entities are not the tangible objects (ill persons) from which they were derived but are abstract concepts whose limits may be revised at will, depending on the particular purpose in mind. In the example just given the entity "fractured tibia" is particularly useful to the orthopedic surgeon in devising therapeutic measures, while the entity "blow by automobile" is more appropriate to the public health worker interested in designing preventive measures.

Descriptive Epidemiologic Observations as Manifestational Criteria

The setting of the limits of a disease entity on the basis of manifestations is often a highly complex process, and an intuitive one. An underlying assumption governing it is that the greater the number and kind of similarities (descriptive of each individual), the more properly the illness of these individuals may be termed an entity; conversely, striking dissimilarities are taken as evidence against the illness being an entity. Attention here is confined to the role played by epidemiologic observations in this process.

Concurrence in respect to characteristics of time or place or person is one of the kinds of similarity commonly used in constructing a disease entity. Numerous examples of this use of descriptive epidemiology may be found. Caverly, describing an outbreak of paralytic disease occurring in Vermont in the summer of 1894, included in his arguments for its being a separate disease

(poliomyelitis) the season of the year during which the illnesses occurred; the aggregation of colic at Devonshire helped to delineate a special type of colic from colic in general (Baker, 1767); and the alternation of anencephaly and spina bifida within the same sibship, plus the fact that both manifestations are noted more among females than among males, provided a reasonable basis for considering them as one entity rather than two (MacMahon *et al.*, 1953).

Epidemiologic dissimilarity, on the other hand, may be used as an argument in favor of dividing one entity into two. For example, Fogh-Andersen (1942) has advanced the opinion that harelip (with or without cleft palate) is an entity to be distinguished from cleft palate (without harelip), because the former is primarily an affection of males and the latter of females.

A Special Class of Experiential Criterion and an Analogy

A special class of experiential criterion, internalization of microorganism, has been so valuable to both treatment and prevention that it deserves study as a model which might have relevance to other classes of experiential criterion. There is no obvious reason, for example, why such another class might not be internalization of social stress.

Five aspects of the application of internalization of microorganism as an experiential criterion appear pertinent to generalizations about the use of experiential criteria in delineating disease entities.

(1) The disease entity so derived bears by definition a *sine qua non* relationship to the criterion as a result of the criterion having been selected (by man, for his own purposes) from an infinite causal complex.

(2) Evidence indicating causal association between internalization of microorganism and morbid processes, although required, does not necessarily have to be watertight. A series of conditions which it would be desirable to fulfill in order to establish such causal connection and frequently referred to as Koch's postulates are itemized by Topley and Wilson (1936) as follows:

(a) The microorganism should be found in all cases of the disease in question, and its distribution in the body should be in accordance with the lesions observed.

(b) The microorganism should be cultivated outside the body of the host, in pure culture, for several generations.

(c) The organism so isolated should reproduce the disease in other susceptible animals.

The comments of Topley and Wilson on technical difficulties which may hinder fulfilling all Koch's postulates in every instance are illuminating. In practice not infrequently the evidence obtained appears to be incomplete by Koch's standards. Further, Koch's third postulate, if indeed it refers to "other susceptible animals" as stated by Topley and Wilson, in reality hedges about establishment of causal association. Man in fact would be the animal necessary for a more convincing demonstration of causal connection between his disease and a microorganism.

(3) The phrase in Koch's first postulate, "the disease in question," implies that *an* existing manifestational entity be necessarily suspected of having causal relationship with a microorganism. Possibly this is true in some instances. However, extraordinary revisions of the limits of previously acceptable disease entities have commonly followed identification of microorganisms

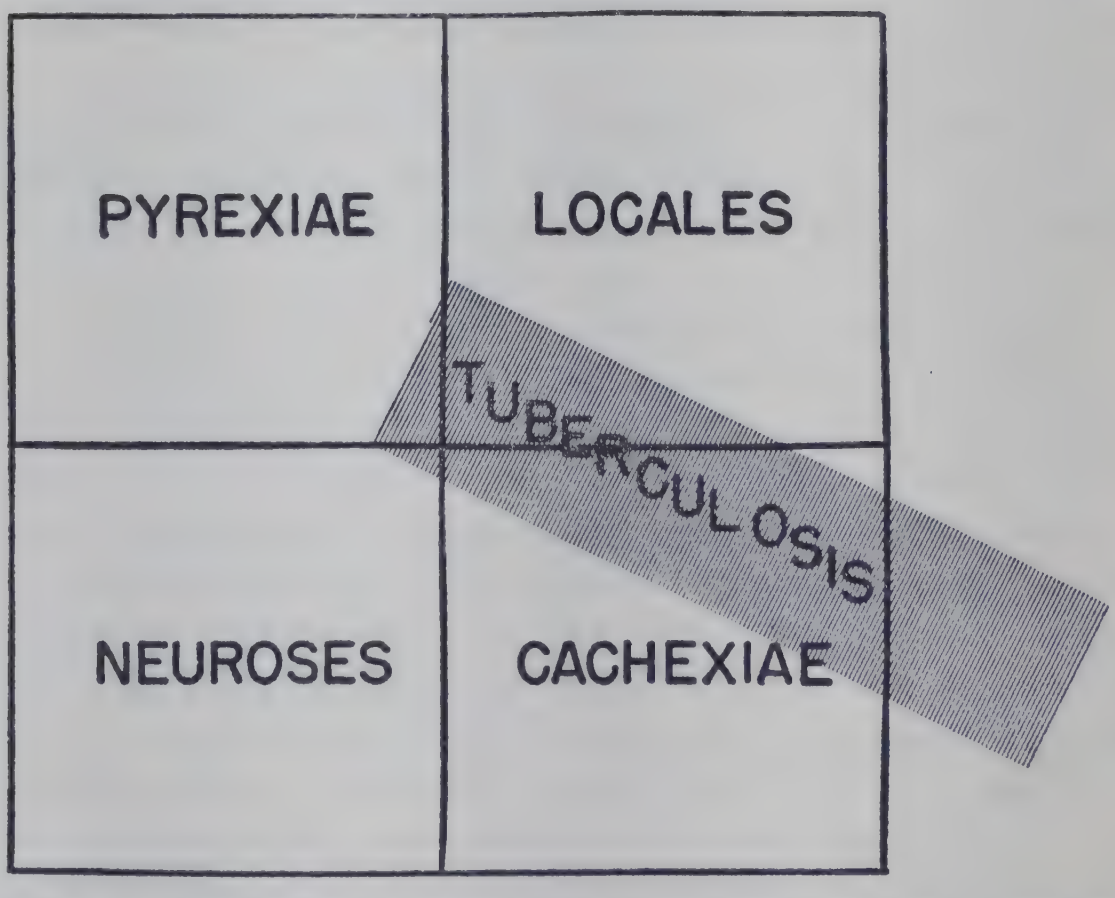


FIGURE 4. Classification of persons with tuberculosis according to manifestational (unshaded) and experiential (shaded) criteria.

which have deleterious effects on man. The revisions which followed identification of the tubercle bacillus by Koch (1882) offer a classic example. The disease entity constructed around this specified experiential criterion, internalization of the tubercle bacillus, removed parts from many manifestational entities found widely scattered among the major categories of disease in the nosologies of preceding times. For example, Cullen's classification (1785) consisted of the four manifestational categories indicated in Figure 4. Each of the four no doubt contained, in differing proportions, ill persons whom today it is found useful to classify as tuberculous (shaded portion of the diagram). It would appear that the limits of the "disease in question" may be determined with clarity only *after* the experiential criterion has been specifically defined.

(4) Evidence has *not* been required that internalization of microorganism be invariably disabling. In Figure 5 an array of ill persons is presented as in Figure 3. Here, however, the model includes non-ill persons in the surrounding area. It is seen that non-ill as well as ill persons may satisfy the selected experiential criterion. The size of the non-ill group may be larger than that of the ill. For example, we presently acknowledge that only a fraction of all persons who internalize the tubercle bacillus become ill as a consequence.

(5) The "necessary" experience may be an experiential

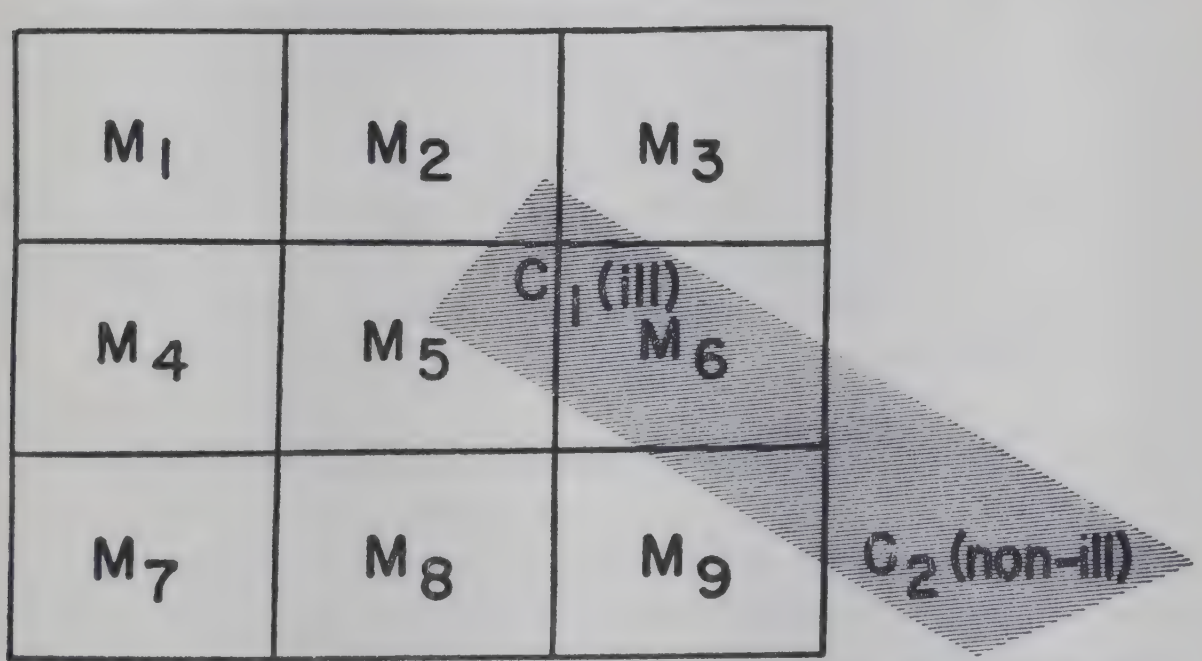


FIGURE 5. Model for classification of ill and non-ill persons according to manifestational (unshaded) and experiential (shaded) criteria.

complex. Two types of microorganism may be required. Swine influenza, for example, depends on the simultaneous presence of bacillus and virus (Shope, 1931). Or the requirement may embrace a non-microorganismal component in conjunction with a microorganismal one. In the guinea pig intratracheal introduction of silica suspension and dead tubercle bacilli produces massive nodular lesions resembling those of silicotuberculosis in man, while neither agent alone excites obvious disease reaction in the experimental animal (Kettle, 1934).

Mention was made at the outset of this section that classes of experiential criterion other than internalization of microorganism might be useful in delineating experiential disease entities, and that another such class might be internalization of social stress. A hypothetical example is mother-death during infancy. It may be noted that this hypothetical criterion is eminently identifiable in all cases, perhaps with a precision greater than that with which microorganisms may be identified. The possibility that such a class of experience may become internalized in a manner analogous to internalization of microorganism has been pointed out by Wiedorn (1955). The "disease in question," as finally delineated, might include parts of such manifestational entities as mental retardation, schizophrenic reactions, personality disorders, etc., just as parts of the pyrexiae, locales, neuroses and cachexiae of Cullen became tuberculosis. Distribution of the experience "in the body . . . in accordance with the lesions observed" is entirely compatible with current ideas that an experience of the nature exemplified is stored in the brain and that the brain is the mediator of perception, thinking, emotion and behavior. Evidence suggesting causal relationship between loss of parent and morbid processes has been demonstrated by numerous case history studies, but how often the experience is disabling has yet to be determined.

Some Postulates for Constructing an Experiential Disease Entity

In connection with the introduction of experiential classifications in place of existing manifestational categories, the following seem pertinent:

(1) The manifestations must be of such a nature that causal

association with the specified experience seems probable, even though the manifestations may seem diverse in that they embrace portions of several accepted manifestational entities.

(2) The experience must be regularly followed by undesirable effects on health in a substantial proportion of those undergoing it and the undesirable effects must not be outweighed by desirable ones.

(3) The experiential entity constructed must have promise of utility.

In what respects must construction of an experiential disease entity have promise of utility? When the experience becomes internalized and continues in time within the recipient, the advantages to the therapist should be obvious. The group of ill individuals at hand have illnesses highly dependent on a common experience, and the experience (whether internalization of a particular microorganism or of stress from mother-death during infancy of the patient) becomes a focus for special therapeutic attention. Antibiotics or specific psychotherapeutic techniques may be developed. In addition, there are advantages to the public health worker interested in prevention. Means may be devised to prevent exposure of the infant to the microorganism or the infant may be provided with fortification such that an inevitable internalization of the microorganism has a minimum of ill effects; similarly, means may be devised, if not to prevent loss of mother, to give infants so deprived supportive measures sufficient to reduce substantially the likelihood of subsequent ill effects.

Finally, lest the impression be left that experiential classifications have some virtue not shared by manifestational ones, it should again be noted that the same group of ill persons may be arranged to create entities of disease according to either manifestations or causal experiences or both, depending on the particular purpose in mind. The word "entity" has come to have such a deceptively comforting sound in current medical usage that J. S. Mill's remarks (1862) seem as pertinent to the medical profession as to the logicians: "Entity, a piece of barbarous Latin, invented by the schoolmen to be used as an abstract name, in which class its grammatical form would seem to place it; but being seized by logicians in distress to stop a leak in their terminology, it has ever since been used as a concrete name."

4

The Strategy of Epidemiology

The predominant purpose of epidemiology was described in Chapters 1 and 2 as the search for causal association between diseases or other biologic processes and specific environmental experiences. The methods used to further this purpose may be classified as follows:

(1) Descriptive epidemiology — description of the distribution disease, with comparisons of its frequency in different populations and in different segments of the same population.

(2) Formulation of hypotheses — tentative theories designed to explain the observed distribution of the disease in terms of causal associations of the most direct nature possible.

(3) Analytic epidemiology — observational studies designed specifically to examine the hypotheses developed as a result of the descriptive study.

(4) Experimental epidemiology — experimental studies on human populations to test in a stringent manner those hypotheses that stand the test of observational analytic studies.

If epidemiology operated independently of knowledge gained in other biologic and medical disciplines, these methods of investigation might be expected to proceed in the order listed. However, according to the contribution made by other disciplines, particularly by clinical medicine and laboratory experiment, any

one, none, or all of these methods may be used in epidemiologic investigations of disease.

The investigation of certain diseases has followed these steps in the order listed. An example is the investigation of cholera. The descriptive phase of the investigation resulted in observations that cholera spread along the lines of human communication, that its progress was never faster than the speed of human travel, that persons coming into close contact with cholera patients frequently contracted the disease, that its occurrence was connected with poverty and overcrowding, and that it was associated with refuse and filth. These facts, plus clinical and pathologic observations, led Snow (1855) and others to the hypothesis that the disease occurred when a healthy person swallowed specific cellular agents discharged from the bowel of a cholera patient. The series of analytic studies that Snow conducted between 1848 and 1854 to test this hypothesis are classics of epidemiologic investigation and will be referred to frequently. The experimental phase of the cholera investigations consisted of attempts to keep drinking water free of contamination by sewage; in spite of the haphazard nature of these experiments and the absence of controlled studies in the modern sense, it was demonstrated that if these efforts were successful, freedom from cholera resulted.

The investigation of the association between regular use of drinking water containing greater than average concentrations of fluoride and low rates of dental caries has also followed this sequence — from early demonstration of geographic patterns of dental caries and of association of mottled enamel with low caries rates to later experiments involving the addition of sodium fluoride to the drinking water of selected populations.

In other areas the gamut of investigation has not yet been fully run. For example, descriptive observations of lung cancer revealed two striking features: (1) the disease is several times more common in men than in women, and (2) its increase during the present century has been striking. These features considered in connection with clinical and pathologic knowledge led to the hypothesis that cigarette smoking is causally related to lung cancer. Many analytic studies designed to test this hypothesis have uniformly supported it. The nature of the environmental influence implicated in this instance is such that an experimental test

has not been conducted in man, and in fact may not be practicable. However, the results of the observational analytic studies are so compelling that there is little question as to the outcome of such an experiment, were it to be conducted.

For many other diseases the hypotheses that have been developed as a result of substantial descriptive data have not received adequate analytic tests. Hypotheses linking cancer of the cervix and circumcision of the marital partner, cancer of the esophagus and alcohol, cerebral palsy and birth trauma, and heart disease and diet may be considered in this category.

In other diseases specific hypotheses have not yet been developed, either because of difficulty in assembling descriptive data on a disease, or because the descriptive variables that are ordinarily measured are not related to the causal factors involved.

DESCRIPTIVE EPIDEMIOLOGY

A logical inquiry into the nature of an unknown quantity must open with questions of a general nature. This is true whether the unknown quantity is a specific environmental situation associated with development of a particular disease or is an unseen object to be identified by a maximum of twenty questions in a parlor game. Thus, in the latter case, when items of great rarity, such as the mummified left great toe of Ramses II, must be identified, the chance of success through unguided guesses is remote. More profitably, the inquiry opens with general limiting questions; for example, is it animal, vegetable or mineral? If satisfactory answers are received to these questions (and in the present example they may not be), the inquiry may proceed with similar general questions as to its age, color, consistency, and so on. Eventually, the number of possibilities compatible with the answers resulting from these general questions may be sufficiently limited for specific guesses.

Like experts in the parlor game, epidemiologists have developed their own set of general limiting questions. The epidemiologist is limited because his questions are not so easily answered as those in the parlor game; consequently, the value of an answer to a general question must be weighed against the expense and time involved in deriving it. Since the collection of data is expensive,

the descriptive phase of epidemiologic investigation is heavily influenced by what information is available, and the variables most frequently examined are those on which information is already at hand in such sources as vital records, hospital records, and published reports.

The variables most commonly examined have been classified as descriptive of *time, place* and *person.* They include:

(1) Characteristics describing the *time* in which persons were found affected:

Year
Season
Day

(2) Characteristics describing the *place* in which persons were found affected:

Country
State, or district within countries
Urban or rural
Local community

(3) Characteristics describing the *persons* affected:

Age, sex and ethnic group
Occupation, education, and socio-economic status
Marital status
Family history

Many descriptive variables do not fit neatly into this classification. For example, order of birth and maternal age might be considered as properly belonging to all three — person, place and time.

The variables listed are those things that are measured, and are not necessarily those which have direct causal connection to the disease being examined. Thus an association of a disease with a particular place might well be a circumstance of the particular persons that inhabit that place, and an association with a variable that is descriptive of a person, for example, his race, might equally well be the result of factors peculiar to the place of residence of that race. The classification should be thought of as a form of mnemonic rather than as an attempt at meaningful categorization.

FORMULATION OF HYPOTHESES

In epidemiology, as in science in general, the thought processes involved in the formulation of hypotheses have not been clearly formalized and classified. Yet a new and convincing hypothesis can be one of the most powerful forces influencing the direction of future research, and the success or failure of that research frequently depends on the soundness of the hypothesis.

The most common hypotheses in epidemiology seek to identify causes sufficient to explain the distribution of a disease noted in its descriptive study. Four methods of reaching hypotheses are discussed.*

The Method of Difference

If the frequency of a disease is markedly different under two different circumstances and some factor can be identified in one circumstance that is absent in the other, this factor, or its absence, may be a cause of the disease.

The difficulty with this method of formulating epidemiologic hypotheses is usually not one of being unable to identify a factor that is present in one circumstance and not in the other (or relatively so), but rather of the multiplicity of hypotheses which may often be suggested on the basis of the difference. For example, literally thousands of possibilities may be suggested to explain differences in disease frequency between males and females, or between the United States and India, or between certain age groups. For this reason, any one descriptive observation, even though revealing a marked characteristic of the disease under study, is seldom sufficient for the establishment of a sound hypothesis. Occasionally one feature of difference between two circumstances is sufficiently striking to favor it as a likely basis for hypothesis, but this situation is unusual. For example, although carcinoma of the cervix is one of the commonest forms of cancer in females, it is extremely rare in nuns. The differences between nuns and other females are not so numerous as to make unworthy of

* The first three methods are patterned after three of Mill's five canons of induction (1862), which dealt with the definition and proof of causation, rather than the formulation of hypotheses.

further investigation the hypothesis that some aspect of the reproductive process predisposes to carcinoma of the cervix.

The Method of Agreement

If a factor is common to a number of different circumstances that have been found to be positively associated with the disease under study, this factor may be a cause of the disease.

For example, cholera is associated with patient contact, with overcrowding, and with sewage-contaminated water supply. A common factor is the opportunity to ingest fecal discharges from cholera patients. This may be a cause of cholera. Again, lung cancer is more common among men than women and was more common in 1950 than in 1900. Cigarette smoking, among other things, shares the same distribution and consequently may be a cause of lung cancer.

The Method of Concomitant Variation

This method involves the search for some mechanism whose frequency or strength varies with the frequency of the disease. It is a quantitative rather than a dichotomous way of looking at the same ideas that are involved in the methods of agreement and difference. A well-known example is the attempt to relate the relative frequencies of various dietary constituents to the incidence of coronary heart disease in different areas of the world. Another is the comparison of the fluoride content of drinking water in various areas to the frequency of dental caries in the inhabitants.

The Method of Analogy

The distribution of a disease may be sufficiently similar to that of some other disease that has been more completely and successfully investigated to suggest that certain causes may be common to the two diseases. This method of arriving at hypotheses is related to the process of deduction, whereby derived epidemiologic principles are applied to specific situations. Thus we tend to associate diseases occurring in certain places during the summer months with vector transmission, and we are reluctant to accept diseases as infectious that do not spread rapidly. Even in the absence of general laws or principles, however, this method may be of service. For example, the observation that the age and

sex pattern for pulmonary tuberculosis in adults had features similar to those of pulmonary cancer and the knowledge that lung cancer is related to smoking led Lowe (1956) to test, with some success, the hypothesis that smoking may also be a factor in the breakdown of pulmonary tuberculosis in later life.

The method of analogy, while useful, is also treacherous; considerable errors have been made as the result of false analogies. Thus, because genetically determined diseases tend to run in families, there has been a tendency until quite recently to assume that familial concentration of a disease is of itself sufficient evidence of genetic determination. This view is of course demonstrably false. Similarly, too close analogy from experimental production of congenital malformations to hypotheses concerning their etiology in man has led to many hypotheses of doubtful value.

Evaluation of Hypotheses

Having discussed the most common methods of arriving at epidemiologic hypotheses, we may note some principles concerning them, based on consideration of those hypotheses that have been most successful in the past.

(1) The value of a hypothesis is inversely related to the number of acceptable alternatives. The number of acceptable alternatives again depends on a variety of circumstances:

(a) In the use of the methods of agreement or difference, the greater the number of circumstances in which the suspected factor is present or absent and in which the disease frequency is correspondingly increased or decreased, the fewer the number of acceptable alternatives. As noted, an association with one variable does not usually lead to the development of a convincing hypothesis. Association with two independent variables may narrow the field considerably. For example, relatively few hypotheses explain both the high incidence of cancer of cervix in married women and the low incidence in nuns, although either one of these associations might have multiple explanations. When yet a third observation is added — for example, the low incidence of cervical cancer among Jewish women — the possibilities are reduced even further.

(b) If two variables are both found to be associated with

a disease, the closer the association between the two variables themselves, the less valuable is the information regarding their association with the disease. For example, many factors may be conceived that are common to one of the Census Bureau's "professional" occupations and residence in a favored residential district. Any disease associated with one of these variables will almost always be associated with the other. Much more limited in number are factors common to holding a professional occupation and, say, singleness, since the correlation between occupation and marital status is weak.

(c) Associations with certain variables, for example, occupation and religion, may be of greater value than those with others, such as age and sex, since the specific environmental circumstances associated with the former may be fewer than those associated with the latter.

(2) It is not necessary that the suspected mechanism be present in all the circumstances under which the disease frequency is found increased. Inconsistencies between what is actually found and the expectation of what will be found based on a hypothesis may be due to:

(a) Multiple causes. As discussed in Chapter 2, our concern is with *a* cause of a disease, not *the* cause. Causes other than the one suspected may be responsible for those observations that do not fit into the hypothesis. For example, not all people who ingest cholera-contaminated sewage develop cholera. This in no way detracts from the hypothesis linking sewage and cholera, since it is quite certain that there are differences between individuals in the degree to which they are susceptible to the organism.

(b) Crudity of the disease classification used. For example, nonsmokers develop lung cancer (although much less often than smokers). At first sight this is inconsistent with the hypothesis linking smoking and lung cancer. However, subsequent investigations have shown that certain uncommon pathologic varieties of lung cancer are not related to cigarette smoking.

(c) Errors of observation.

(3) New hypotheses are commonly formed by relating observations from several different fields. In the area of disease causation, epidemiologic findings are most profitably viewed in the light of clinical, pathologic and laboratory observations. Thus

Snow interpreted the epidemiologic observations of cholera in the light of clinical and pathologic observations that the intestine was the apparent focus of the disease.

(4) The stronger a statistical association, the more useful is it as a basis for a hypothesis. "Strength" means not the degree of statistical significance, but the degree to which the situation of a disease being entirely absent in one circumstance and invariably present in another is approached. Associations of low strength, however significant statistically, are rarely productive of strong hypotheses.

(5) Change in the frequency of a specific disease over time, although very difficult to substantiate, is one of the variables that is most productive of hypotheses. This is particularly true of changes that occur over a relatively short period. For example, in London in the early nineteenth century, the water was supplied by a number of different private companies. Snow (1855) observed that in 1849 death rates from cholera were high in those areas in which the water was supplied by the Southwark and Vauxhall Company and the Lambeth Company. Five years later, in 1854, the number of deaths in the area supplied by the Southwark and Vauxhall Company was somewhat higher, but there was a substantial re-

TABLE 1

Mortality from cholera in the areas of London supplied by the Southwark and Vauxhall, and Lambeth Water Companies in 1849 and 1854 *

Districts with water supplied by	Number of deaths attributed to cholera	
	1849	1854
Southwark and Vauxhall Company	2261	2458
Lambeth Company	162	37
Both companies	3905	2547

* From Snow, 1855.

duction of deaths in the Lambeth Company area (Table 1). There was no large scale population shift that might have accounted for the change. It was attributed by Snow to the fact that in the intervening years the Lambeth Company had changed its sources of supply to the higher reaches of the Thames River, above, Snow supposed, the greatest sources of contamination.

The attractive feature of this hypothesis is the relatively few

factors that can be found to have undergone rapid changes coincident with the observed change in disease frequency.

(6) An isolated or unusual case should receive particular attention. Such cases, for example, in persons living in communities otherwise free of the disease, have been of great value in the infectious diseases and figure prominently in the investigations of Snow on cholera and in those of Panum on measles (1846).

ANALYTIC EPIDEMIOLOGY

The testing of a hypothesis may in part be carried out by further descriptive observations — observations that were not available at the time the hypothesis was formulated. For example, if a relationship between smoking and lung cancer is hypothesized on the basis of the unusual sex ratio and the increasing incidence over time, one could further inquire whether lung cancer is most frequent in those countries where cigarette smoking is most prevalent, or whether lung cancer is relatively infrequent among occupational, ethnic or religious groups whose members smoke little or not at all. Wynder and Lemon (1958) examined the incidence of various cancers and of coronary heart disease among a group of Seventh-day Adventists. The group experienced only 10 per cent of the expected incidence of lung cancer. Successful prediction of associations based on a hypothesis is usually considered stronger evidence in favor of the hypothesis than is knowledge of the same associations at the time of the formulation of the hypothesis. However, this type of evidence is indirect, and is usually capable of alternative explanations. More direct methods of testing the hypothesis are usually sought.

The most satisfactory evidence is that needed to test arguments of the type "If the difference between two populations in disease frequency is dependent on differences between the populations with respect to a specific factor, then, within each of the populations, this factor should be found more frequently in people with the disease than in those without." For example, if the difference between males and females in the frequency of lung cancer is the result of heavier smoking by males compared to females, then higher rates of lung cancer should be found among males and females who smoke than among those who do not.

The data required to test such an inference are rarely available in the common sources from which descriptive observations are made, and special studies directed specifically to obtaining the data are usually required. A classic example of such a study is Snow's investigations of cholera in London. As already noted, Snow formed the hypothesis that the frequency of cholera in various geographic sections of London was related to the sources of water supply. He was particularly interested in the Southwark and Vauxhall Company which drew its water from the section of the Thames River most heavily contaminated with sewage, and the Lambeth Company which in 1854 drew water from the river above the reaches of greatest contamination. As shown in Table 2,

TABLE 2

Mortality from cholera in the districts of London supplied by the Southwark and Vauxhall, and the Lambeth Companies, July 8 to August 26, 1854 *

Districts with water supplied by	Population 1851	Deaths from cholera	Cholera death rate per 1000 population
Southwark and Vauxhall Company	167,654	844	5.0
Lambeth Company	19,133	18	0.9
Both companies	300,149	652	2.2

* From Snow, 1855.

the areas of London supplied entirely by the Southwark and Vauxhall Company experienced in the summer of 1854 a rate of 5.0 deaths from cholera per 1000 population, whereas the death rate in the areas supplied by the Lambeth Company alone was only 0.9 per 1000. A large area supplied by both companies experienced a rate of 2.2 deaths per 1000, a rate midway between those for the areas supplied by either of the two companies alone. Clearly these observations are consistent with the hypothesis that persons drinking water supplied by the Southwark and Vauxhall Company had greater risks of cholera than those drinking Lambeth Company water. However, there may be many factors other than their water supply that differed between these geographic areas in a way similar to that in which the cholera rates varied.

A more critical test of this hypothesis is provided by an examination of the data for the area in which the water of both companies was available. If the water supply were related to the risk of cholera death, then, within the area supplied by both companies, those persons supplied by the Southwark and Vauxhall Company should have a higher death rate than those supplied by the Lambeth Company. This test would be particularly critical because, as Snow pointed out, within this area the two companies supplied persons of similar economic, occupational and age status. Frequently, adjacent houses were supplied by different companies, and often the persons supplied were themselves unaware of the source of their supply.

However, while information was available on the number of houses supplied by each company, no data were recorded on the source of water used by each cholera patient; consequently, special investigation was required. Snow inquired of relatives and others as to which company supplied water to each house in which a death from cholera occurred between July 8 and August 26, 1854. The results are shown in Table 3. Within the area sup-

TABLE 3

*Mortality from cholera in London, July 8 to August 26, 1854, related to source of individual water supply in three groups of districts**

Group of districts with water supplied by	Water supply of individual houses	Population 1851	Deaths from cholera	Cholera death-rate per 1000 population
Southwark and Vauxhall Company	Southwark and Vauxhall Company	167,654†	738	4.4
Lambeth Company	Lambeth Company	19,133†	4	0.2
Both companies	Southwark and Vauxhall Company	98,862	419	4.2
	Lambeth Company	154,615	80	0.5
Rest of London	—	1,921,972	1,422	0.7

* From Snow, 1855.

† Overestimated by a small amount, since this figure includes population with no water supply.

plied by both companies, the cholera death rates for customers of each company were very similar to those of the same company's customers in the areas supplied exclusively by that company.

Moreover, Snow pointed out, the death rate for customers of the Lambeth Company was no higher than that for the rest of London, even though the majority of the Lambeth Company's customers were located in the area supplied also by the Southwark and Vauxhall Company — an area in which the epidemic raged severely. The hypothesis that the drinking of water supplied by the Southwark and Vauxhall Company was associated with death from cholera was therefore confirmed.

Of course, there is no clear-cut dividing line between descriptive and analytic epidemiology. Many studies contain elements of each, and in many cases there is often great difficulty in deciding whether a "specific" hypothesis is being tested or not. For example, examination of the geographic distribution of a disease might be considered a test of the hypothesis that the disease does or does not exhibit geographic variation in incidence. Nevertheless, in practice, most studies can be categorized as being directed primarily toward either formulation or testing of rather specific hypotheses.

Types of Analytic Study

To illustrate the general forms that analytic epidemiologic studies may take, let us suppose that a study is to be undertaken to determine whether an association exists between maternal infection with rubella and congenital malformations in the offspring. With respect to the two variables (rubella and malformation) every individual must fall into one of the cells in the following fourfold table, the frequencies being indicated by the letters *a, b, c* and *d:*

Maternal rubella during pregnancy	Malformation in offspring		Total
	Present	Absent	
Present	a	b	$a+b$
Absent	c	d	$c+d$
Total	$a+c$	$b+d$	$a+b+c+d$

The question at issue is whether malformation is more common in the offspring of mothers who have had rubella during

pregnancy than in the offspring of mothers who have not. In other words, one wishes to compare:

$$\frac{a}{a+b} \text{ with } \frac{c}{c+d}$$

There are two general ways in which such a comparison can be made. The most direct would be to take two groups of women, one of which $(a + b)$ had and one of which $(c + d)$ did not have rubella during pregnancy, and determine how many of each group showed malformation in the offspring. This method is based on the primary selection of an exposed and an unexposed population, and it provides a direct measure of the risk of being affected according to whether the suspected causal factor was present or absent.

A second method would be to investigate the frequency with which rubella had occurred among the mothers of malformed and not malformed children. In this method the primary selection is of groups of affected $(a + c)$ and unaffected $(b + d)$ persons, with subsequent inquiry as to the frequency of the suspected cause in the two groups. This enables a comparison of:

$$\frac{a}{a+c} \text{ with } \frac{b}{b+d}$$

If $\frac{a}{a+c}$ is greater than $\frac{b}{b+d}$, then $\frac{a}{a+b}$ must be greater than $\frac{c}{c+d}$. However, except in unusual circumstances (see Chapter 14), the values $(a + b)$ and $(c + d)$ are not determined in this type of study, and direct estimates are not obtained of the *extent* to which the risk of bearing malformed offspring among women exposed to rubella during pregnancy exceeds that of nonexposed women.

These two methods of investigation are commonly described as prospective (primary sampling of exposed and not exposed) and retrospective (primary sampling of affected and not affected), respectively. The rationale of this terminology is that the retrospective method involves "looking backward" from effects to preceding causes and the prospective method involves "looking forward" from supposed causes to subsequent effects. White and Bailar (1956) have pointed out that this terminology gives rise to

considerable confusion, since the terms retrospective and prospective have everyday meanings quite different from those implied in their application to these special methods of study. In the general sense, a retrospective study is one based on past data or past events, whereas a prospective study is one planned to observe events that have not yet occurred. To some extent there is overlap between the general and epidemiologic meanings of these terms. Thus, retrospective epidemiologic studies are usually, though not always, retrospective in the general sense. However, prospective epidemiologic studies may depend on data already accumulated or on the occurrence of events subsequent to the primary sampling of the population. The method of prospective investigation using past data is an economical and satisfactory one and is being used increasingly. For example, Case and Lea (1955), to examine in 1955 the relationship between mustard gas poisoning and lung cancer, selected a group of men who in 1930 were receiving pensions for the effects of mustard gas poisoning received during World War I, and examined the mortality experience of this group between 1930 and 1952.

In order to avoid confusion of terms we will refer to studies in which the primary selection is of affected and nonaffected persons as "case history studies," and to studies in which the primary selection is of exposed and nonexposed groups as "cohort studies." "Cohort studies," a term already in common use, indicates the investigation over time of an identified group or cohort of individuals. These terms are, in our opinion, clear and descriptive of the methods involved.

The two methods have quite distinct areas of usefulness as well as their individual difficulties, which are discussed in some detail in Chapter 13 (Cohort Studies) and Chapter 14 (Case History Studies). It may be useful here, however, to compare these uses and difficulties in general terms.

Cohort studies have two major advantages over case history studies:

(1) Direct estimates of the risk associated with the suspected causal factor are obtained. Such estimates may sometimes be obtainable from case history studies, but the estimates will not be direct and will only be possible if the ratio of affected to unaffected individuals in the population is known. This ratio is not

usually known in a case history study, since a sample of the unaffected population rather than the total is commonly used and the sampling fraction is often unknown.

(2) Spurious relationships resulting from bias in data-collecting procedures are somewhat less likely in cohort than in case history studies. For example, in case history studies, data on the frequency of past exposure to the suspected cause are frequently based on the patients' memories of such exposures. Yet affected individuals (e.g., the mothers of malformed infants) may be more highly stimulated to remember past experiences than more fortunate individuals. Even when more reliable sources of information are used, the possibility of more thorough investigation of the affected than of the unaffected individuals is a constant problem in case history studies. In cohort studies, on the other hand, the division of the population into exposed and unexposed groups may be made without knowledge of, or even commonly prior to, the appearance of the end result, thus reducing considerably the possibility of this type of error.

Cohort studies have one considerable disadvantage. By comparison with case history studies they are generally laborious, time-consuming and expensive. Related to these characteristics is the additional difficulty of assembling in one study sufficient data for conclusive answers. This is illustrated in the studies that have been undertaken of the association of maternal rubella with congenital malformation. Even the combined results of four cohort studies (Hill *et al.,* 1958) give only general ideas about the magnitude of the risks involved.

The reciprocal advantages of case history studies are that they are relatively quick and inexpensive, they are easily repeatable, and large numbers of cases can be examined economically. Therefore, case history studies are commonly used when it is desirable to test as rapidly and as cheaply as possible several of the hypotheses that may have been offered to explain observed relationships, and to determine initially which of the several hypotheses are worthy of further exploration. Cohort studies are more frequently undertaken when the evidence from case history studies is fairly conclusive that a relationship exists, in an effort to obtain more detailed and more reliable data. This pattern has been followed in a number of investigations in recent years. In the investigations of

the relationship between smoking and lung cancer, for example, some fifteen case history studies had been published prior to the appearance of the first cohort study in 1954 (Doll and Hill, 1954).

These generalizations about the uses of the two methods of study may be modified by a number of characteristics of the disease under study. For example:

(1) The more common the disease or trait to be investigated, the easier will be a cohort study, since the smaller will be the population that need be surveyed to identify a given number of cases. Conversely, the rarer the disease, the more difficult becomes a cohort study, and the more desirable a case history study.

(2) The shorter the interval between the suspected cause and the end result in a particular individual, the easier will be a cohort study. For example, cohort studies of association between events in pregnancy and characteristics of the offspring that are manifest at birth are greatly facilitated by the relatively short follow-up interval needed to record outcome. Much of the objection to cohort studies is therefore removed. Conversely, study of the relationship between maternal deprivation in infancy and mental illness in adult life would be extremely difficult to attempt by the cohort method.

(3) The more objectively the history of the suspected cause can be recorded, the more satisfactory will be the case history method.

(4) A strong statistical association between the suspected cause and the end result will make either kind of study easier, but because of the relatively greater expense of a cohort study (due to the necessity of assembling data on large numbers of unaffected individuals), the benefits of such a strong association will be more apparent in this type of study. For example, although only 83 cases of leukemia were recorded during 12 years of observation of the Hiroshima City population exposed to the atomic bomb (Wald, 1958), the relationship between degree of exposure and leukemia risk was so strong that it could be demonstrated unequivocally even on these numbers. (The leukemia rate for the most heavily exposed persons was about sixty times that for those lightly exposed.) Obviously, because of the even larger numbers

required, it is difficult to substantiate weak associations by means of cohort studies.

EXPERIMENTAL EPIDEMIOLOGY

Up to this point we have been discussing methods of determining whether a statistical association exists between a certain environmental situation and a certain disease. To a variable extent inference may be possible as to whether a demonstrated association is causal or not. However, as indicated in Chapter 2, proof of the causal nature of an association depends on experiment — the demonstration that alteration of the effect follows alteration of the reputed cause.

Because of the extensive writings of Greenwood, Topley and Wilson (see, for example, Greenwood *et al.*, 1936), the expression "experimental epidemiology" has been used to describe studies of the mechanisms of spread of infectious diseases in laboratory animals. However, the meaning of the expression may be broadened to include experimental studies of hypotheses of disease causation in man.

In a way, the application of any public health measure that is based on knowledge of disease etiology may be considered an experiment, since it is a test of the validity of that knowledge. Examples of such experiments range from the trials of various foods as antiscorbutics during the seventeenth and eighteenth centuries to the chronic disease control programs that are under trial at the present time. Epidemiologic observations are usually required to test the success of such measures.

A more limited concept of experimental epidemiology is that it comprises studies or programs *specifically* designed and operated to test etiologic hypotheses. Such studies differ from programs that are set up primarily for the purpose of providing service in that control or comparison groups are usually provided for in the former but not in the latter. Whenever feasible it is desirable that experimental programs precede service programs. When this does not happen, there is the danger that unrelated events following the service program may be misinterpreted as evidence for or against the validity of the knowledge on which it

is based. A downward trend in disease incidence may be wrongly interpreted as a consequence of the service program when it was due to other coincident changes, or a stationary incidence may be wrongly interpreted as evidence of the lack of effect of the service program if other factors were productive of increasing incidence during the same period. Once a therapeutic or preventive practice has achieved acceptance, whether on sound or unsound grounds, it is very difficult to introduce experimental tests if indicated at a later date.

Such considerations have led to a widening use of formal experimental designs in the preliminary stages of both therapeutic and preventive procedures. In the preventive field, such studies have been designed to test the efficacy of various immunizing agents in the prevention of infectious diseases, to test the association of dietary deficiencies with specific disorders, and, most recently, to test the causal nature of the association between dental caries and fluoride-deficient water supplies.

5

Measurements of Disease Frequency

Statements of the frequency of a given trait or disease manifestation in various populations are essential to descriptive epidemiology. Such statements permit comparisons between populations and between subgroups of a population with respect to the manifestation in question. In this chapter, the several measures of frequency in present use are described. Our concern is with the definition of these measures and with the theoretical concepts underlying them. Practical considerations that limit their accuracy are discussed in later chapters.

In simplest terms a statement of frequency might be, "There were 500 cases of tuberculosis," the only requirement being a count of patients with tuberculosis. However, the statement has little utility until it is further qualified with respect to two essential features: (1) in what population and (2) when were these cases observed. Thus, the statement would gain needed precision if it read, "On January 1, 1960, in the population of this town there were 500 cases of tuberculosis." For administrative purposes, such a measure of frequency might be entirely satisfactory. For example, information in this form would facilitate the planning of medical facilities for the tuberculous patient in that particular town, since such planning basically requires only knowledge of the number of cases and not necessarily of the probability of contracting tuberculosis.

MP 100 N-60
0541(
COMMUNITY HEALTH
LIBRARY
AND
DOCUMENTATION

If, however, the purpose of the frequency statement is to determine whether one town contains mechanisms of greater or less strength in the production of tuberculosis than another, such a statement of frequency has obvious shortcomings. The second town may have more or fewer cases than the first because its population is larger or smaller. To allow for differences in population size, the frequencies must be expressed in the form of *rates.*

RATES

A rate is a "quantity or degree of a thing measured per unit of something else" (Webster). In epidemiologic usage the "thing measured" is the trait or disease manifestation while the "unit of something else" is the population (diseased and healthy) to which the observed cases belong. The time at or during which the cases (trait or disease) and population are studied is a further specification which is essential for epidemiologic purposes. Thus, the tuberculosis frequency rate in a town on January 1, 1960 might be expressed as 500 cases per two million population (the actual population of the town). In practice, to enable rapid comparison with other rates, rates are expressed not in relation to total populations but in relation to certain convenient units of population size — usually some power of ten. The tuberculosis rate in the town might then be expressed as .025 per cent, .25 per thousand, or 250 per million. Such rates are derived by dividing the number of cases by the actual population and multiplying by the chosen unit of size.

It is evident, therefore, that three items of information are necessary in epidemiology for the expression of any rate — the numerator of the fraction (the number of persons affected), the denominator (the population among whom the affected persons are observed), and a specification of time. The denominator is commonly called the *related* or *reference* population. When the denominator is restricted solely to those persons who are capable of having or contracting the disease, it may properly be considered as a population *at risk;* the rates or probability statements thus derived are of particular pertinence to an understanding of etiology.

Both the numerator and denominator of a rate must be de-

fined in similar terms; if the numerator is restricted to certain age, sex or racial groups, the denominator must be similarly restricted, and vice versa. For example, if the denominator population is the number of residents of a town, cases appearing in the numerator should also be limited to residents of that town. Similarly, if the numerator consists of the number of cases of a disease limited to females (e.g., uterine disease), the denominator would necessarily be the related female population.

The denominator of a rate may not be a "population" in the ordinary demographic sense. For example, a hospital may express its maternal mortality rate as so much per thousand deliveries. Although the women delivered do not form a "population" in the general sense, they may still be referred to as the related population and in this case as a population at risk, since the statement derived describes the probability of death among a defined group at risk. Similarly, the case fatality rate is the number of deaths from that disease per so many patients with the disease; here the diseased group constitutes the population at risk.

Ratios. Epidemiologic rates are in fact statements of probability, in the statistical sense. Certain measures, which in general usage might be called rates, are not probability statements, and are often called ratios. One common form of ratio is the expression of the number of affected persons as a proportion of the number of unaffected persons (rather than as a proportion of the affected plus unaffected persons). For example in the United States it is common to express the number of fetal deaths as a proportion of the number of related livebirths. This is called the fetal death ratio and is distinguished from the fetal death rate which is the number of fetal deaths as a proportion of the related total births (livebirths plus fetal deaths). Ratios, although satisfactory for simple descriptive purposes, do not share with rates the advantage of being immediately amenable to treatment as probability statistics.

DISEASE ACCOUNTING

The amount of disease existing at a given point in time, like the level of a bank account at the end of a month, depends on the size and number of prior credit and debit transactions. Un-

like a bank balance, unfortunately, the disease balance is never found to be negative. But procedures similar to those descriptive of bank accounts are applicable in explaining disease levels, the difference being limited to the terms used in naming the amount on hand and the various forms of transaction (addition and removal).

Point Prevalence (Amount on Hand)

Point prevalence of a disease is its frequency at a given moment in time. It is a measure of what prevails or exists, and corresponds to the current bank balance. Expressed as a rate for a specified population, point prevalence is the proportion of that population who exhibit the disease at a particular instant. The numerator includes all persons having the disease at the given moment, irrespective of the length of time elapsed from the beginning of the illness to the time when the point prevalence is measured. The denominator is the population among whom the disease is ascertained. Although prevalence is sometimes stated in the form of period* prevalence, our use of the word generally is restricted to point prevalence.

Incidence (Addition)

The amount of disease on hand at a given moment (point prevalence) is the result of an interplay of dynamic forces. One of these forces is the *incidence* — the frequency of addition, or transaction *into* the disease account. In order to specify the frequency of such additions it is necessary to indicate the time during which cases are added.

The incidence rate of a disease is defined as the number of cases of a disease that come into being during a specified period among a specified unit of population.

In practice it is often impossible to measure incidence directly, since definite timing of an onset of illness is commonly uncertain. Usually certain arbitrary points in the disease process are used as approximations. Thus, measurements such as onset of symptoms,

* The number of cases that were observed *during* (rather than *at*) a specified time, for example during a year. Period prevalence, therefore, consists of the point prevalence at the beginning of a specified period of time plus all cases that arise during the period. Period prevalence is a particularly complex measure, since both point prevalence and incidence are incorporated in it.

time of diagnosis, and date of notification or of hospitalization are judged to be events related to how many cases arise within a specified period of time. Incidence rates, it should be noted, are measures of the frequency of *events,* as distinct from the *existences* measured by prevalence rates.

There are only two methods by which additions may be made to the disease account of a population: by the onset of illness (either an initial illness or a recurrence), and by immigration of persons already ill into the population.

Recovery or Death (Removal)

The other dynamic force determining point prevalence is the removal *from* the disease account. There are three general ways by which the disease account may be depleted. For many diseases death is the most frequent means of leaving the population. A second frequent way is by recovery, either from the initial or a recurrent attack. The third way, emigration of persons with the disease, either by observed emigration from the population or by loss to observation, may be of importance in illnesses of long duration.

Death. It is necessary to distinguish two rates descriptive of death: (1) the disease-specific *death* or *mortality rate,* and (2) the disease-specific *case fatality rate.* In both the numerator is the same — the number of persons dying of the disease during a certain period (see page 68 for discussion of why it is sometimes unnecessary to specify a period of time for case fatality rates). The denominators of the rates, however, are different. In a mortality rate the denominator is the total population within which the deaths occurred, while in a case fatality rate it is restricted to persons with the disease. Thus the mortality rate expresses the probability that a member of a defined *total* population will die of the disease, while the case fatality rate expresses the probability that a member of a population of *affected* individuals will die of the disease. The former rate is most useful to epidemiology since it is related, through the case fatality rate, to the incidence of the disease; the latter is most useful in describing the clinical course of the disease once it is established and in evaluating the efficacy of various treatments directed against it.

Recovery. Ordinarily, the term *recovery rate* relates the number

of recoveries to the number of persons ill from the disease. It is an uncommon practice to relate the number of recoveries from a disease to the total population.

Cause of death. In a disease-accounting procedure any individual known to die during the observation period must be withdrawn from the account. Such individuals may be divided into two groups — those who die *because* of the disease and those who die *with* the disease from some other cause. Although distinction between these groups is unimportant for the basic accounting procedure, it becomes necessary to separate the two if determination of the numerator of mortality and case fatality rates is desired, since the concern in theory is with the first group only — those who die because of the disease. Two methods are used to ascertain this group:

(1) Judgment, based on review of individual cases. For example, a woman with cancer of the cervix with bladder involvement who dies in uremic coma without evidence of other pathologic states may be said to have died of cervical cancer, whereas had the same woman been killed in an automobile accident she would have been said to have died from an accident while suffering from cervical cancer. For such a judgment to be valid as an index of mortality attributable to cervical cancer requires that women with cervical cancer experience no greater mortality from accidents than do women in general. Such an assumption is of doubtful validity, even for the example chosen, and is still less certain for such associated conditions as diabetes and heart disease, nutritional defect and infection, and so on. On theoretical grounds, the method is unsound since it involves making judgments about causal association in individual events. As noted in Chapter 2, valid inferences concerning causal association can only be made for categories of events of which the individual events are merely examples.

(2) Statistical assessment, based on the subtraction from the total number of affected individuals who die of that number who might have been expected to die if the general death rate among nonaffected individuals of similar age and sex had obtained. Although theoretically the superior method, this procedure requires complete ascertainment of the number of affected individuals at risk of death and of all who die, whatever the alleged cause. Al-

though rarely practicable for deriving mortality rates, this method often does have applicability in estimating attributable case fatality. The statistical considerations in this method have been described by Muench (1959).

An important function of death data in epidemiology is to provide crude estimates of incidence or prevalence when direct estimates cannot be made. When death data are put to such use, it should be borne in mind that most national and other vital data on mortality rates from specific diseases are dependent on judgments concerning individual cases as described under (1), page 56. Since 1948 in the United States and 1940 in Great Britain, deaths have been tabulated according to the "underlying cause of death" as selected by the certifying physician from the various pathologic states that might be present at the time of death. While grossly inadequate with respect to certain conditions (e.g., diabetes), the statistics gained in this way yield highly accurate measures of mortality from certain others (e.g., acute leukemia). Mortality rates so derived are most satisfactory as indices of either incidence or prevalence of conditions which (a) are likely to be selected as the underlying cause if present at the time of death, and (b) run a sufficiently short course that the probability of affected persons dying in conjunction with other reported causes will be small. Death from a disease is a particularly close index of incidence (as contrasted to prevalence) when the disease is uniformly fatal. On the other hand, when the probability of a person having died from one particular cause is high, the frequency of other manifestations present at the time of death approximates prevalence. A recent example is the use of autopsy material from soldiers killed in combat to estimate the prevalence of atherosclerosis (Enos *et al.*, 1955).

If the method of statistical assessment is used to derive the numerator for an attributable case fatality rate, the number of deaths obtained will at times be substantially larger than the number which would have been obtained through use of the method of individual judgments. This disparity will be least for diseases which run a short course and greatest for chronic diseases, where, in fact, deaths from "other" causes may far exceed those which attending physicians would have specified as having direct connection with the disease in question.

Example of Measures of Frequency

To illustrate some of the measures of frequency, use will be made of the data shown in Tables 4 and 5 on cases of leukemia that are known to have existed in the white population of Brooklyn, New York, between 1948 and 1952. For the present purpose, we are assuming that these represent all the cases of leukemia in the white population of the city, which in 1950 numbered 2,525,000. The tables show the accounting of cases for each year in the period. The total cases existent in each year (column 3) are derived from those patients alive at the beginning of the year who had received a diagnosis of leukemia in a previous year (column 1), plus cases newly detected in that year (column 2). Columns

TABLE 4

*Data on 417 patients with acute leukemia, Brooklyn, 1948-1952**

Year	(1) Diagnosed in a previous year	(2) New cases diagnosed in year	(3) Total cases existent in year	(4) Deaths in year	(5) Patients lost trace of in year	(6) Patients alive at end of year
1948	7	69	76	54	7	15
1949	15	91	106	86	3	17
1950	17	83	100	73	3	24
1951	24	99	123	101	1	21
1952	21	68	89	81	1	7
Total	84	410	494	395	15	84

* There were 7 patients alive at the beginning of the period whose acute leukemia had been diagnosed previously. Diagnosis of 410 new cases was made during the period, for a total of 417 cases known to have existed during the period. Of this total 395 patients died during the study period, 15 were lost trace of during the period, and 7 were still alive at the end of the period.

4, 5 and 6 show the disposition during the year of the cases in column 3. The number of patients known to be alive at the end of the year (column 6), is, of course, the same as the number alive at the beginning of the next year (column 1).

Prevalence. Patients alive at the beginning of the year whose disease was diagnosed in a previous year (column 1) are in fact the cases of leukemia existent at 00.00 A.M. on January 1 of each year. They form, therefore, estimates of point prevalence. Point prev-

TABLE 5

*Data on 495 patients with chronic leukemia, Brooklyn, 1948-1952**

Year	(1) Diagnosed in a previous year	(2) New cases diagnosed in year	(3) Total cases existent in year	(4) Deaths in year	(5) Patients lost trace of in year	(6) Patients alive at end of year
1948	129	79	208	50	8	150
1949	150	83	233	71	5	157
1950	157	90	247	90	6	151
1951	151	61	212	84	7	121
1952	121	53	174	89	4	81
Total	708	366	1074	384	30	660

* There were 129 patients with chronic leukemia alive at the beginning of the survey period whose disease had been diagnosed in previous years. Diagnosis of 366 new cases was made during the period, for a total of 495 cases known to have existed during the period. Of this total, 384 patients died during the study period, 30 were lost trace of, and 81 were still alive at the end of the period.

alence rates can be derived by relating these numbers to the population at risk. For acute leukemia the average point prevalence for these five points in time is:

$$\frac{84}{5} \times \frac{10^6}{2{,}525{,}000} = 6.7 \text{ per million.}$$

Column 3 shows the numbers of cases existent during each year. Average annual period prevalence rates can therefore be derived. For acute leukemia, the rate is:

$$\frac{494}{5} \times \frac{10^6}{2{,}525{,}000} = 39.1 \text{ per million.}$$

An important characteristic of period prevalence rates is evident in these data. The period prevalence does not bear a direct relationship to the length of the period of observation. For example, the 5 year period prevalence ($7 + 410 = 417$) is less than five times the average annual period prevalence ($494/5 = 98.8$). Clearly some patients who would be included only once in a 5 year prevalence would be included several times in successive annual period prevalences.

Incidence. Time of diagnosis is used here as the basis for determining incidence. Incidence rates are derived from column 2. The 5 year incidence rate for acute leukemia is:

$$\frac{410 \times 10^6}{2{,}525{,}000} = 162.4 \text{ million per 5 years.}$$

Since incidence rates are usually expressed per annum, this may be divided by 5 to derive an average annual incidence rate of 32.4 per million per year.

Mortality rate. The mortality rate is derived from column 4. For acute leukemia the average annual death rate is:

$$\frac{395}{2{,}525{,}000} \times \frac{10^6}{5} = 31.2 \text{ per million per year.}$$

This estimate is based on the assumption that all deaths of affected persons were caused by leukemia. Actually, on the basis of the age-specific death rates from all causes in the general population at the time, one of the patients with acute leukemia and 18 of those with chronic leukemia might have been expected to die from other causes during the 5 year period. The subtraction of these cases from the total deaths makes little difference to the mortality rate, and the assumption seems to be a reasonable one for this particular disease.

The various rates described in this section are summarized for both forms of leukemia in Table 6.

TABLE 6

Frequency rates for acute and chronic leukemia derived from the data in Table 4 and Table 5

Average annual rates per million population	Acute leukemia	Chronic leukemia
Incidence	32.4	29.0
Death (mortality)	31.2	30.4
Period prevalence	39.1	85.1
Point prevalence (not annual)	6.7	56.1

INTERRELATION OF INCIDENCE AND PREVALENCE

There is an important relationship between prevalence and incidence which may be expressed as follows:

P_t varies as the product of $I_{(t)}$ and d

where P_t is point prevalence at the end of the period (t), I_t is inci-

dence during time (t), and d is the average duration of the disease from onset* to termination.

It follows that a change in point prevalence from one period to the next may be the result of changes (1) in incidence, (2) in duration, or (3) in both incidence and duration. For example, improvements in therapy, by preventing death but at the same time not producing recovery, may give rise to the apparently paradoxical effect of an increase in prevalence of the disease. And decrease in prevalence may result not only from decrease in incidence but also from a shortening of the duration of illness — through either more rapid recovery or more rapid death.

If incidence and duration are constant over a succession of time periods the disease is said to be in equilibrium and:

$$P_{t+1} = P_t$$

In this circumstance the relationship between prevalence, incidence and duration is such that:

$$P_t = I_{(t)}d.$$

For a disease in equilibrium, this formula may be used to calculate the other component of the equation when two are known. For example, from the estimates of prevalence and incidence of leukemia in Table 6 the average duration of the disease can be calculated from the relationship $d = \frac{P_t}{I_{(t)}}$:

$$\text{acute leukemia: } d = \frac{6.7}{32.4} = 0.21 \text{ years} = 2.5 \text{ months.}$$

$$\text{chronic leukemia: } d = \frac{56.1}{29.0} = 1.93 \text{ years} = 23 \text{ months.}$$

These are close to the values of 2.4 months for acute leukemia and 20 months for chronic leukemia derived independently from special follow-up studies of the same patients. These durations represent the interval between diagnosis and death, since incidence here is based on date of diagnosis and death is the only avenue of removal from this disease account (other than emigra-

* If incidence is measured from some stage other than onset (e.g., diagnosis), duration is measured from the same stage. Prevalence must also include only cases existent at or subsequent to the same point in the disease process.

tion with disease). Note that the formula gives duration in the same time unit as that used for incidence (in this case, years).

The above formula is most useful in consideration of diseases in which the predominating form of removal is either recovery or death. It must be remembered that duration to death in some diseases is shorter, and in others longer, than duration to recovery.

Another useful relationship is that between incidence, $I_{(t)}$, and mortality, $M_{(t)}$, which may be used to measure the case fatality rate, f:

$$f = \frac{M_{(t)}}{I_{(t)}}$$

Thus, for acute leukemia (see Table 6) $f = 31.2/32.4 = 0.96$, and, for chronic leukemia, $f = 30.4/29.0 = 1.05$. The findings are consistent with the general conception that both forms of leukemia are invariably fatal. Calculations dependent on this relationship should again be restricted to instances in which (1) the disease is in equilibrium, and (2) the reports of death and incidence are equally complete. If the disease is not in equilibrium, but incidence is either increasing or decreasing from one time to the next, the ratio mortality/incidence will be lower or higher, respectively, than the case fatality. Such a discrepancy becomes larger with increasing lengths of time from onset to death.

MEASUREMENTS PARTICULARLY RELEVANT TO EPIDEMIOLOGY

A variety of rates have been described which are related to the frequency of a disease in a population. The use of one rather than another is dependent on the relevance of the one selected to the particular problem in mind. Where the problem is epidemiologic the rate of choice obviously is incidence.

Incidence

Attempts at elucidation of causal factors are concerned with an event or occurrence — *becoming* ill — and the measure most directly descriptive of this event is the incidence rate of the disease. Causal factors necessarily operate prior to the "onset" of disease. Therefore, the closer the stage of disease at which incidence is

measured comes to the time of actual onset, the more directly "incidence" will be influenced by the operation of causal factors. Incidence, potentially, is the most direct measure for comparing populations with respect to causal factors.

Death Rates as Indices of Incidence

The main reason why death rates are so often used as indices of incidence is their ready availability. Routine registration of deaths together with specification of "cause" of death enables death rates to be derived for many diseases and for many populations for which more direct measures of incidence are not available. Note should again be taken that death rates reflect the characteristics of incidence rates most closely for a particular disease when the interval between onset and death is short and when the disease in question has a high case fatality rate.

Prevalence

For epidemiologic purposes point prevalence measurements are inferior to incidence measurements, since two sets of factors determine point prevalence: those connected with the occurrence of cases, and those connected with course and duration (prognosis) of the cases once they are established. Although the same factors may at times be operative in both mechanisms, there would appear to be a much greater chance of correctly assigning the role of individual factors if the two processes were studied separately by investigation of incidence rates on the one hand, and of prognosis on the other. When substantial incidence data are scarce, point prevalence data at times give clues as to incidence, but the contribution of duration to point prevalence must be constantly borne in mind.

On the other hand, many administrative situations require knowledge of how many patients with a given disease exist in the community. Clearly the answer to such a question rests in point prevalence. Incidence rates are of uncertain value since, for a given incidence rate, diseases of long duration impose a greater community burden than those of short duration. Even here, however, incidence rates may be of considerable importance if, as is often the case, facilities for new cases differ from those for patients who do not recover within a short period of time.

Although period prevalence is less useful epidemiologically than other measures of disease frequency, it is given attention here because of its recurring use in an epidemiologic context. Period prevalence is a compound of (1) a past point prevalence (the resultant in turn of a previous interrelation between incidence and duration) and (2) a more recent incidence. It is therefore influenced by variation from a large number of factors. Furthermore, because there is no direct relationship between period prevalences measured over different periods, period prevalence is a most difficult measure to evaluate.

Theoretically, period prevalence could be used to estimate incidence or point prevalence if one or the other of these were known, $P_{(t)} = P_t + I_{(t)}$. But in practice if data are available to estimate period prevalence as well as one of the other measures, they are usually available for direct calculation of the third measure.

Occasionally, period prevalence data are used as estimates of incidence or point prevalence. The extent to which this is justifiable depends on, among other things, the following considerations:

(1) The longer the period over which period prevalence is determined, the closer period prevalence resembles incidence.

(2) The shorter the period over which period prevalence is determined, the closer period prevalence resembles point prevalence.

(3) The shorter the average duration of disease, the closer period prevalence resembles incidence.

(4) The longer the average duration of disease the closer period prevalence resembles point prevalence.

SPECIFICATION OF TIME

We have seen that an incidence rate measures the frequency of events in a certain sized population over a specified period of time. In specifying the period over which the events are observed, two things must be defined: (1) the point in time at which the observation began, and (2) the length of the period of observation. In the case of point prevalence rates no duration of observation is involved, but it is still necessary to specify the point in time at which the observations are made.

Points in Time

So far we have spoken of time as if calendar time were necessarily implied. Actually there are several ways of specifying time.

One way is by the specification of a calendar time. If, for example, we wished to estimate the incidence of coronary thrombosis in men aged 45, we could observe all cases of coronary thrombosis in men aged 45 with onset between January 1 and December 31, 1958, and relate this to the population of 45-year-old men at the middle of the year, to derive an annual incidence rate. Since the date of occurrence of events is very frequently recorded and it is consequently relatively simple to decide whether or not an event occurred between two calendar dates, this is the most common way of specifying the beginning and end of a period of observation.

Similarly, it is common to measure point prevalence at a point in calendar time. National censuses, for example, which measure the point prevalence of people, are taken as of a specified date.

A second way to specify time is by using chronologic age of the population. In the previous example of coronary thrombosis in 45-year-old men, an alternative method would be to observe a population of men who had just attained their forty-fifth birthday. Since not all men reach their forty-fifth birthday at the same time, the periods of observation would not all coincide in calendar time.

A common and more practical use of chronologic age is in a situation in which there is interest in a disease occurring within a short time of birth. For example, in a maternity hospital over a certain period of time there may have been 2000 deliveries of liveborn children, 22 of whom died in the first week after birth. The death rate in the first week of life may then be expressed as 11 per 1000 livebirths. Not all these children were observed during the same period in calendar time, and it does not matter for this purpose whether the 2000 births occurred over a period of 1 year or of 10 years. The period of observation is specified as beginning at a certain age (birth) and continuing for 7 days thereafter.

In the case of point prevalence data, an example of a similar type of measurement might be a statement that by age 10 years

20 per cent of a certain population are tuberculin positive. Most commonly, such a measurement would be defined in terms of both age and calendar time. Thus the statement might read "In 1959, at age 10 years, 20 per cent of this population are tuberculin positive." However, derivation of a point prevalence requires specification of a *point* in time, not a period as vague as a year, and it is neither practical nor desirable to specify this point in terms of *both* variables (i.e., measuring the tuberculin sensitivity of persons all reaching their tenth birthday at the same point in time). For purposes of assembling the data, the specification may be in terms either of age or of calendar time. Thus according to the type of specification used, the statement might more precisely read "Among persons reaching their 10th birthday at some time in 1959, 20 per cent were tuberculin positive," or "Among persons who were ten years old on July 1, 1959, 20 per cent were tuberculin positive." The implications of these two statements are for all practical purposes the same, but the method of assembling the data is quite different.

Occasionally this type of rate is implied in the way a rate is expressed, although the actual method of derivation of the rate may be on the basis of a specified calendar time. Thus, in the previous neonatal mortality example (see page 65), the hospital may have been able to observe each of the infants born for a period of 7 days and to record deaths occurring in the actual denominator population. In the case of the much more frequently used *infant mortality rate,* however, this may not be possible, since this would necessitate observation of the denominator population for one year. It is usual therefore to express the infant mortality rate as:

$$\frac{\text{Number of deaths of infants aged less than one year}}{\text{Number of livebirths during the same year}} \times 1000$$

The implicit assumption is that the number of livebirths during the period of observation is the source of all the infant deaths occurring during the same period. Since infant deaths are heavily concentrated in the early part of the first year of life, this is not an unreasonable assumption, particularly when the annual number of births is relatively constant.

In theory the relationship $P_t = I_{(t)}d$ holds whether age or calendar time is the point of reference. However, the same proviso holds for age as for calendar time: the disease must be in equilibrium. In the situation where age is the point of reference this means that neither incidence nor duration changes appreciably with age. In fact we know of no disease for which this is true; changes in incidence with age are usually so marked as to make calculations from the formula of little value. The general principle, however, that point prevalence is a function of incidence and duration remains the same, and point prevalence at a specified age depends on:

(1) Incidence at younger ages.
(2) Duration. This in turn depends on:
    (a) Continuance of the observed state.
    (b) Mortality of persons manifesting the state relative to those not manifesting it.

An example of this principle is observation of characteristics evident at birth, such as congenital malformations. Although such observations are frequently expressed as incidence rates (e.g., 1.3 cases of harelip per 1000 births), they are in fact point prevalence rates, since they are observations of the frequency of existence of certain states at a certain age (birth). The events which might be used to express true incidence rates (the occurrence of harelip) occurred at least 6 months before birth. What such rates are called is of less concern than an understanding that they have the characteristics of point prevalence rates in being simultaneously influenced by both incidence and duration. In this case, since natural recovery from harelip does not occur, the most important determinant of duration is survival. Differences in the frequency with which congenitally malformed fetuses die and are aborted will produce differences in the point prevalence of congenital malformations at birth, even though the incidence of malformation is the same.

A third method of specifying the time of onset of an observation is according to the appearance of some characteristic or the occurrence of some experience that is common to a group of persons but without clear specification of either age or chronologic time. Thus, we speak of the incidence of complications of a certain disease — the proportion of patients with the disease who ex-

perience complications. Another example is the incidence of complications of pregnancy. Case fatality rates are similar types of statements. None of these rates bear much relationship to rates used for describing disease frequency in the population at large.

The prevalence of defects among military inductees or among school entrants forms a special class. Here there is a tacit specification of age, and the frequency of defects noted may, in some circumstances, be considered representative of the population from which the entrants were drawn.

Lengths of Time

An incidence rate involves specification of the period of time during which events have occurred. The use of arbitrary periods such as weeks, months or years enables ready comparison of data from several studies.

However, in measurement of incidence rates of some diseases or events the population may be at risk for a limited period of time only. In such cases it may be possible to express a "total" incidence that would remain the same regardless of the length of observation. The limitation of the period of risk may be due to the action of etiologic influences that exist for only a short period of calendar time, or to the restriction of risk to certain age groups. For example, this situation frequently exists when incidence is being measured in relation to the number of births (population), since many diseases are limited to infancy. Thus, in Birmingham, England, between 1940 and 1949 there were 194,216 livebirths, and 578 of these infants developed hypertrophic pyloric stenosis (MacMahon *et al.*, 1951). The incidence of pyloric stenosis may be expressed as $578/194{,}216 \times 1000$ or 2.98 per 1000 livebirths without any specification of the duration of observation. Since this condition occurs predominantly in the first 3 months of life and is practically unknown over 6 months of age, a relatively short observation period ensures the derivation of a total incidence, which will not be appreciably increased by increasing the observation period. A similar situation exists with respect to the measurement of the frequency of complications of pregnancy and of specific acute diseases. In the measurement of case fatality rates in chronic diseases it is usually necessary to specify a period during which the case fatality or the frequency

of other complications is being measured. It is common, for example, to speak of 1 year, 5 year or 10 year survival rates.

It is clearly not possible to derive such a "total" incidence rate for diseases which can be contracted at any time in life. An approach to it is the *life time probability* of contracting or dying from a given disease. This is not an observed rate but is a theoretical expectation derived from currently observed incidence or death rates in each age group of the population under the assumption that the currently observed rates will continue to operate throughout the life of the population in question. It is necessary further to assume an arbitrary life span, for example, 70 years, while a true "total" incidence rate for diseases which can be contracted at any time in life would be the probability of contracting or dying from the disease regardless of how long the person lived; as has been noted, this may be determined only when the period of risk is limited to a certain period of time or age span.

A "total" incidence resulting from limitation of risk to a period of calendar time is illustrated by an *attack rate*. This rate contains in the numerator all the events occurring during an arbitrary period of time usually encompassing some circumstance such as the passage of an acute epidemic.

A *secondary attack rate* is a measurement in which the numerator consists of the number of cases of a disease which occur within the same household following the occurrence of the first, or "primary," case. It is usually used in connection with infectious disease, and there is usually a stated or implied time limitation that, on the basis of the known incubation period of the particular disease, indicates that the secondary cases probably derived from the primary case. The denominator in a secondary attack rate is usually the population at risk, rather than the reference population, in that it is restricted to groups actually at risk and, in diseases conferring prolonged immunity, excludes persons who have previously had the disease.

SPECIFICATION OF THE EVENT

The general type of event — onset of illness — which goes into the numerator of incidence rates requires further specification.

If expressed as a rate, incidence (unlike point prevalence) is not necessarily a fraction of the population. The numerator may be made up of cases or attacks rather than individuals, since in some illnesses, such as acute alcoholic delirium or the common cold, some individuals may experience more than one attack during the period of observation. In such a circumstance the probability statement derived measures the risk of cases of the illness occurring among the population rather than the risk of individuals becoming ill. It would seem preferable, for epidemiologic purposes, to reduce the numerator to individuals so that a probability of individuals having one or more attacks during the interval might be expressed. Further subdivision into risks of having 1, 2, 3 . . . attacks during the period are useful in identifying and describing groups of individuals who have an abnormal frequency of attack.

In addition, although attacked only once during the period of interest, some individuals may be experiencing a recurrence of such an illness as reactive depression, their initial attack having occurred several years before the period of observation. Under this condition an incidence rate expresses the risk of individuals having an attack during the period of interest, whereas it would have been more helpful to an understanding of basic causal mechanisms to identify those individuals having their initial episode of the illness during the period of interest. A further reason for avoiding inclusion of persons having a recurrence is particularly applicable to diseases in which recovery occurs in only a fraction of the initial episodes. The characteristics of individuals who have recovered, and therefore are susceptible to a recurrence, are of necessity related to recoverability. Inclusion of such persons in the numerator of an incidence rate would seem likely to becloud the features distinctive to persons having their first episode.

SPECIFICATION OF THE POPULATION

The denominator of incidence rates also presents problems beyond those of identifying its geographic or administrative boundaries that were discussed earlier. If the denominators of incidence rates contain people who are not at risk of contracting the dis-

ease, important epidemiologic differences between the rates may be obscured.

In those acute infectious diseases which confer permanent immunity the majority of a specified population may not be at risk of attack because of their past encounter with the disease. In common chronic or relapsing diseases of extremely long average total duration, the denominator may in turn contain an appreciable proportion of persons who are either currently ill with the disease or have had it in the past and are, therefore, not at risk of having an initial attack.

Ideally an incidence rate, in order to express true "risk," would have its population base confined to those persons who have not in the past had and do not presently have the disease under study, but this is a refinement which has not been commonly practiced, and which in fact may be ignored without serious error unless the proportion of such persons in the population is large.

6

Sources of Mortality and Morbidity Data

An epidemiologist obviously must depend on observations made by many others. One person rarely has the opportunity to observe a sufficiently large population (including both affected and unaffected individuals) to arrive at firm estimates of disease frequency. Implicit in this is the limitation that data constituting the principle sources of information are frequently gathered and published for different purposes than those in which the epidemiologist is interested, and may consequently not suit his purpose completely.

DEATH REGISTRATION

The introduction of death registration was the foundation of modern epidemiology; it turned the subject from a narrative discipline into a science that depends on quantitative arguments for deduction and generalization. For an account of the history of death registration the reader is referred to a recent publication by Dorn (1959a). The primary reasons for recording deaths are of course legal and political.

Deaths reported by cause were the principal reliance of epidemiologic intelligence in early times. They supplanted the less specific total death rate as measures of disease prevalence, al-

though even today changes in the total death rate may give the first indication of epidemic conditions (see Figure 1, page 6). The specification of cause of death has over centuries followed developments in medical knowledge and in concepts of manifestational and experiential classification; international agreements have been established, and more nations are approaching completeness in reporting. "Completeness" means not only that all actual deaths are recorded, but also that assigned causes are shifted from vague and ambiguous categories, such as "marasmus" and "senility," to ones that are as explicit as possible in light of prevailing knowledge.

Death registration contributes importantly to knowledge of the history of certain diseases through secular trends in mortality. In the analysis of such trends, certain difficulties must be considered.

First, it is necessary to take account of changes in the completeness of registration of deaths. In the United States, satisfactory levels of registration have been achieved by the different states at different times. Admission to the "Death Registration Area" — the area for which mortality data are published by the National Office of Vital Statistics — has been dependent upon achievement of satisfactory levels of registration, and consequently the geographic area included in the Death Registration Area has varied from time to time. At the time of its original formation in 1880 the Death Registration Area included only two states — Massachusetts and New Jersey. The Area became complete, that is, comprised all the then existent states, with the inclusion of Texas in 1933. Since death rates vary from state to state, as well as from time to time, these additions to the Registration Area must be considered in the examination of time trends. Trends for the years betweeen 1900 and 1933 are most accurately determinded by restriction of data to the ten registration states which provided acceptable data throughout that period. In England and Wales, long-standing traditions and uniformity of registration procedures are responsible for satisfactory levels of registration at least since the beginning of the present century.

The successive revisions of the *Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death* (the "International List"), an internationally accepted guide for classification and coding of causes of death, may also profoundly

influence the apparent trends of mortality of certain diseases as tabulated in published statistics. This is particularly true with respect to the Sixth Revision in 1948. The identification of changes attributable to such revisions is usually easily made by noting the abruptness of the changes and their coincidence with institution of the revisions. Making allowance for changes due to the revisions is less easy, although there have been special studies of their effects on various causes of death (Van Buren, 1940; Dunn and Shackley, 1944; World Health Organization, 1952).

Another source of confusion results from the changes that have been made in the procedure for classifying death certificates on which the attending physician has noted the presence of more than one pathologic condition. Since practically all published statistics deal with dead persons as if each had but one cause of death, it is necessary to assign each certificate to some one rubric of the International List even though the certificate might logically satisfy criteria for any of several. For many years it was customary to assign such certificates on the basis of certain arbitrary rules established in an internationally accepted *Manual of Joint Cause of Death;* thus, a certificate mentioning both heart disease and diabetes mellitus would automatically have been assigned to diabetes, even though the certifying physician might have considered heart disease the underlying cause of death and diabetes only a coincident condition. Following the recommendations of various international bodies, Great Britain in 1940 and the United States in 1949 began assigning certificates bearing multiple causes on the basis of the certifying physician's opinion as to the "underlying cause" of death. This change had a profound effect on the statistics of some diseases. For example, diabetes mellitus showed an abrupt decline of between 30 and 50 per cent in the number of assigned deaths. Once again, special studies have been made which assist in the interpretation of data that span the years of change (Registrar General, 1947; Erhardt and Weiner, 1950; National Office of Vital Statistics, 1951).

Finally, the formula $M_{(t)} = I_{(t)} \times f$ (see Chapter 5, page 62) indicates that changes in mortality may result from change in either incidence or fatality. This relationship, while contributing to the difficulties in the interpretation of mortality trends, may also occasionally be put to good use. Thus estimates may be made

retrospectively either of incidence from known case fatality or of prevalence from knowledge of both case fatality and duration. For example, before the antibiotic era, the prevalence of tuberculosis in a population could be estimated at ten times its mortality, based on an average 5 years' duration and 50 per cent case fatality for this disease. Currently, this estimate must be modified because of lower case fatality, and an average duration still to be determined.

The improvement in accuracy and completeness in death registration furthers historical studies of disease, but the consequent increased complexity of registration procedures has some drawbacks in terms of immediate usefulness. Processing, classifying, listing, and publishing the contents of death certificates delay the distribution of information on current epidemics. Total numbers of deaths on the other hand can be made available in a relatively short time. The 1952 fog in London, referred to previously (page 5), focused attention on the value of "excess total deaths." The United States Public Health Service, National Office of Vital Statistics, publishes weekly reports of the number of deaths in 114 selected cities. During the influenza pandemic in 1957 these numbers were plotted and compared with the median number of weekly deaths for the preceding 3 years in the same cities (Figure 6); the excess mortality was quite strikingly demonstrated.

Death Reporting in Limited Populations

Official death registration has the main purpose of estimating mortality in the total population. Deaths are also reported in limited portions of the population and these data may be used for specific epidemiologic purposes. For example, some insurance companies have analyzed in detail deaths occurring among their policy holders. Risk of death has been related to attributes such as social class, and physical attributes, such as weight, for which data are not available in national statistics.

Reports of deaths among the Armed Forces under noncombat conditions have greater than usual accuracy and completeness, but naturally pertain to a selected population of limited age and sex range.

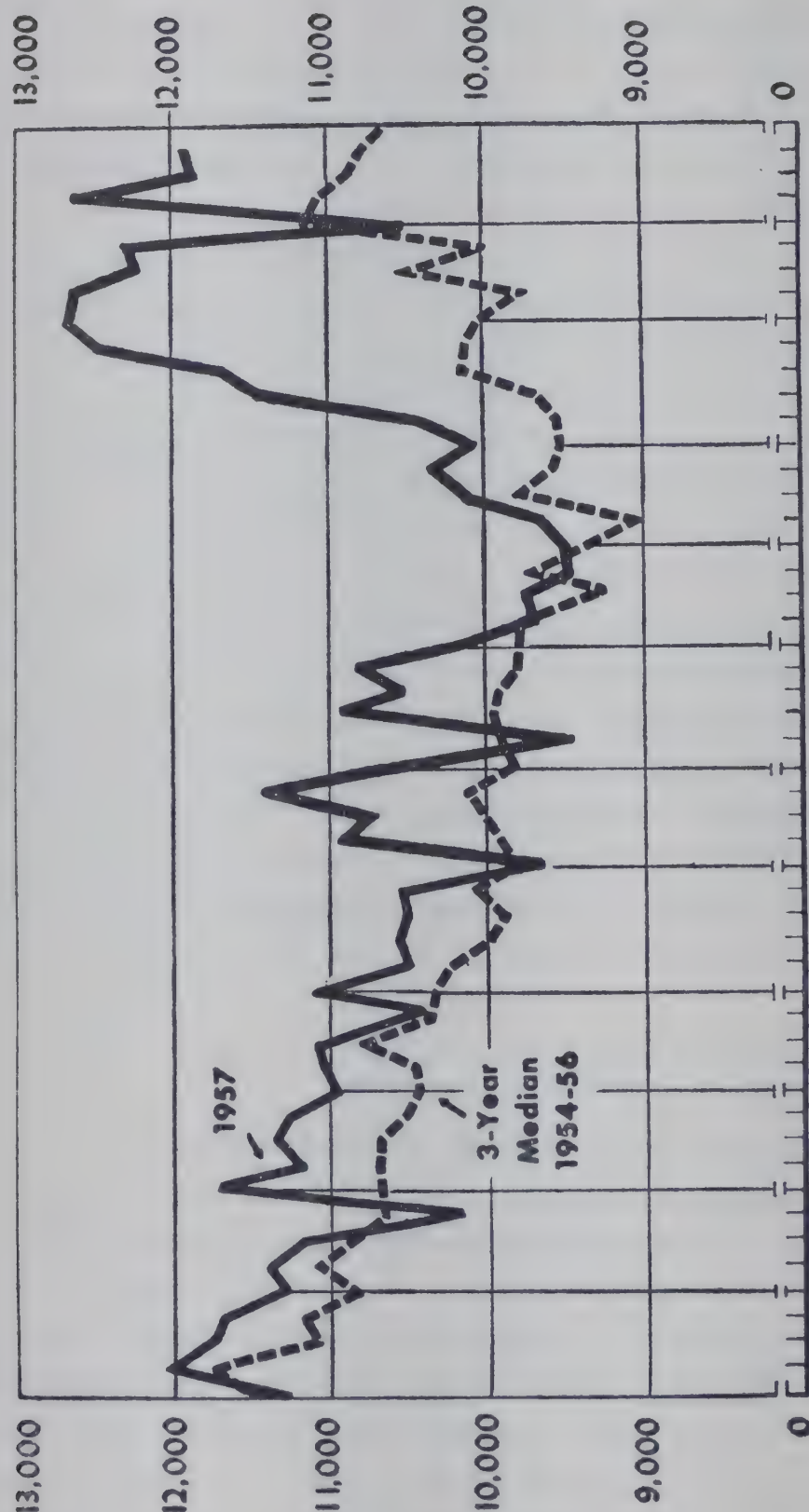


FIGURE 6. Weekly numbers of deaths in 114 selected cities of the United States, 1954 to 1956 and 1957. (Reproduced with permission from the National Office of Vital Statistics, 1957.)

MORBIDITY REPORTING

Official reporting of disease occurrence occasions many more difficulties than death reporting. Apart from certain limited populations, such as the Armed Forces, in which there is distinct administrative interest in morbidity data, the motivation for reporting ordinarily originates from health officials. The routine sources of morbidity data to which the epidemiologist can turn without conducting a special survey include:

(1) Reporting of certain acute diseases to health authorities.
(2) Registers of certain diseases.
(3) Reports on absenteeism from work or school.
(4) Medical statistics of the Armed Forces.
(5) Hospital reports.
(6) Other sources.

Disease Reporting

Legislation requiring the reporting of certain diseases at onset has the objective of maintaining a current picture of the danger which these diseases present to the community. Since a specific diagnosis is required, reporting is usually the responsibility of practicing physicians. Reports are made to local health authorities so that control measures may be initiated with the shortest possible delay. The selection of diseases to be reported is mostly a local matter except for six diseases — cholera, plague, louse-borne relapsing fever, smallpox, louse-borne typhus fever, and yellow fever — the reporting of which is required by International Sanitary Regulation. Reportable (or "notifiable") diseases are usually either infectious diseases or diseases believed to be occupationally determined.

There are two forms of reporting: (1) reporting of cases and (2) reporting of epidemics. Reports of epidemics may be made in the event of unusual outbreaks of a disease, whether or not the disease in question is included on the regular list of reportable diseases, and even, in fact, when the nature of the epidemic is unknown.

Although the purpose of disease reporting is to obtain incidence data, generally the number of cases reported is far fewer

than the total cases occurring. However, valuable inferences can be derived from the observation of relative fluctuations. For example, the trend of reported cases over several years may be assumed to be proportional to the actual incidence trend (providing the period is not too long).

Apart from their use for initiating control measures, the value of reports of epidemics lies mainly in forming a picture of the geographic distribution of disease. For example, during the influenza epidemic of 1957 there was published a weekly map that showed the counties of the United States that had reported epidemic conditions similar to influenza (Communicable Disease Center, 1957). For certain diseases — poliomyelitis, influenza, and arthropod-borne encephalomyelitis — the Communicable Disease Center, United States Public Health Service, issues periodic surveillance reports based on data collected from all over the United States that include many details of unusual completeness.

Disease Registers

Some health authorities attempt to maintain a continuous accounting of patients with diseases of major public health significance, such as pulmonary tuberculosis and cancer, and of patients hospitalized for mental disorders. Newly recognized cases are entered and maintained in the file until recovery, death, or migration from the community. Obviously, it takes several years to build up such a file to completeness. Potentially, these registers are sources of accurate information as to duration of disease, case fatality, incidence, prevalence, and other pertinent epidemiologic measurements. Official reports from such files are often of an administrative nature, dealing with the medical care and welfare assistance to which patients are entitled by local regulations. However, the original data and the reports that are occasionally published in great detail, such as the report of the experience of Connecticut cancer registration between 1935 and 1951 (Griswold *et al.,* 1955), are valuable sources of epidemiologic information.

An example of a particularly complete disease register is the pulmonary tuberculosis register kept by the Central Tuberculosis Dispensary for the City of Copenhagen since early 1930. The main entries of the annual reports are compiled in Table 7 for

the years 1949 through 1956. The table demonstrates some of the purely administrative problems that influence prevalence data, such as the decision to change the criterion for recovery (freedom from *Mycobacterium tuberculosis*) from 5 years to 3 years, and the changes in procedure with respect to the "lost" cases. Com-

TABLE 7

Itemized accounting of prevalence of pulmonary tuberculosis in a City *

Disease entry	1949	1950	1951	1952	1953	1954	1955	1956
(1) Cases on file January 1	7423	7272	6853	6425	4761	4275	3745	3413
(2) New cases during the year	899	796	686	649	572	501	437	372
(3) Relapses after 3 yrs. freedom from *Mycobacterium tuberculosis*	—	—	—	—	59	86	110	72
(4) Persons with tuberculosis entering community	92	119	128	99	101	78	66	69
(5) Deaths from tuberculosis	215	173	173	114	66	71	61	62
(6) Deaths from other causes	24	35	31	48	38	61	59	68
(7) Regarded as recovered	622	734	708	1844†	866	747	633	676
(8) Moved to other communities	238	301	278	279	201	166	163	127
(9) Removed from file because of unknown fate‡	43	91	52	127	47	150	29	38
Point prevalence (1)	7423	7272	6853	6425	4761	4275	3745	3413
Incidence (2) + (3) + (4)	991	915	814	748	732	665	613	513
Termination (5) through (9)	1142	1334	1242	2412	1218	1195	945	971
Total population (in thousands)	764	766	763	760	760	756	752	746

* Source: annual reports from the Central Tuberculosis Dispensary, Copenhagen, Denmark.

† In 1952 the criterion for recovery was changed from 5 years to 3 years freedom from *Mycobacterium tuberculosis*. In the years following 1952 the report contains a separate item for cases relapsed after 3 years freedom from *M. tuberculosis*.

‡ Criterion changed: During the period 1949-1951, meaning no information for 6 years. During the period 1952-1953, meaning no information for 5 years. During the period 1954-1956, meaning no information for 3 years.

parison of lines 5 and 6 indicates the increased fraction of deaths from "other" causes that can be expected when duration of a chronic disease is increased. This information was available because the center was kept informed of all deaths among its registrants as well as of the causes of death that were reported.

Absenteeism

The proportion of persons absent from work or school on any given day is an indication of the total point prevalence of disease and disability in the population. These data are not routinely available as official reports, but under certain circumstances, such as pandemic influenza, reports of total absenteeism can be requested by the health authorities. Excess absenteeism will then yield an important measure of the extent of undue prevalence of disease. Specific causes of absenteeism are usually difficult to determine. A diagnostic statement is rarely available before the worker or the pupil returns, and, even so, may be of questionable accuracy.

Absenteeism may also be used as a measure of the relationship between total population morbidity and certain characteristics of individuals. For example, characteristics such as overweight and physical disability may be related to total number of work days lost in a specified period to determine whether such characteristics are liabilities to health or performance.

Medical Statistics of the Armed Forces

The annual reports of the Surgeons-General of the Armed Forces contain statistics on morbidity. Although these are limited to certain age groups of adult men that are selected because of physical fitness, the information is valuable because of its accuracy.

Hospital Reports

It would be desirable to place hospital records foremost on the list of sources of epidemiologic information. Paradoxically, it appears that official reports from hospitals are the least informative of all sources. Florence Nightingale is credited with the suggestion that all hospitals be required to report admissions by disease category in order that authorities may have dependable and com-

plete information of the disease status of the population. This has never been successfully accomplished, and annual hospital reports are usually restricted to administrative data.

The outstanding exception to this generalization are the data from hospitals for mental disorders. The availability of detailed statistics on mental hospital patients on a local, state and national basis is probably related to most such hospitals being under government control. Statistics are available both in the annual reports of specific hospitals, and, more usefully to the epidemiologist, in central compilations of data for the state or for the country. In some areas, special efforts have been made to include data for hospitals not directly under fiscal control of the agency collecting such statistics. For example, data for Massachusetts now include patients admitted to Veterans Administration and private hospitals as well as the State facilities. Obviously, without such inclusions, major biases of the data with respect to age, sex and socio-economic status are unavoidable. Reports published on the basis of mental hospital data distinguish between census counts of patients under hospital care as of certain dates (prevalence) and first admissions of persons not previously in a mental hospital (incidence). Again these data will be strongly influenced by administrative decisions as to what constitutes being under hospital care (for example, whether this includes all patients on the books, or only those physically in the hospital), what is a "first" admission, and what constitutes a "discharge" from the hospital.

Other Sources

Particular local circumstances sometimes enable estimates of morbidity from specific diseases. For example, the allocation of priority supplies of certain foods on the basis of medical certification of illness during the rationing of food in Great Britain has enabled estimates of the prevalence of diabetes, thyrotoxicosis, tuberculosis, peptic ulcer and other conditions according to age, sex and geographic location (Stocks, 1949).

MORBIDITY SURVEYS

The routine sources of morbidity data discussed previously have mostly been designed to estimate prevalence or incidence by

striving for a complete disease census of the population studied. In addition, periodic attempts have been made to supplement these sources by special surveys of samples of the population. For descriptive purposes, these special surveys can be grouped in two general classes:

(1) Studies of the frequency of a specific disease or group of diseases in the population of a limited geographic area believed to be generally representative of a larger population.

(2) Special surveys of defined populations with the attempt to estimate morbidity from all causes.

Perhaps the best-known examples of the first group are the surveys of cancer morbidity in ten metropolitan areas of the United States conducted by the U. S. Public Health Service in 1937 and 1947 (Dorn and Cutler, 1959). A similar survey has been conducted for the State of Iowa (Haenszel *et al.*, 1956), and individual investigators have published many reports of less extensive surveys, usually covering a smaller geographic area and limited to a narrower group of diseases. The technique of such surveys is usually to assemble reports of all known cases of the disease in question from practicing physicians, hospitals and death records. The reports must of course be sifted to exclude persons who are not residents of the area under investigation and to avoid duplication of reports of the same patient from different sources. The variety of sources included and their completeness will determine to a considerable extent the accuracy of the survey. In a survey of the prevalence of multiple sclerosis in the City of Boston between 1939 and 1948, 62 per cent of the cases were identified from hospital records, an additional 34 per cent were ascertained by inquiry of practicing physicians, and 4 per cent, which were not recorded by hospitals or physicians, were traced from death certificates (Ipsen, 1950). These proportions will, of course, vary with time and place, and with the disease under investigation. Thus, in the surveys of cancer in the ten metropolitan areas considerable variation between sites of cancer was noted in the percentage of cases reported by physicians (as opposed to hospitals): from 51 per cent for skin cancer to 15 per cent for cancers of the urinary system and of the central nervous system.

Surveys of those diseases in which it is felt that an appreciable proportion of affected individuals do not come to medical atten-

tion often require a special procedure. With such diseases door-to-door interviewing may be conducted to ascertain previously unknown cases. This procedure is particularly pertinent to studies in the area of mental disease, and a number of such investigations have been conducted in recent years (Opler, 1958; Leighton, 1959). For chronic diseases prevalence, but not incidence, is usually determined by such surveys.

The other class of morbidity survey, in which total morbidity is to be ascertained, involves the selection of representative samples of the population. Since the information required can only be obtained directly from individuals constituting the population, surveys of the total population would be prohibitively expensive. Generally, information has been obtained by household interview, supplemented sometimes by the use of diaries written by the families involved. Two surveys undertaken by the Commission on Chronic Illness, one in Baltimore (1957) and one in Hunterdon County, New Jersey (1959), were unusual in that medical examinations were also offered to the persons interviewed.

Much valuable epidemiologic data as well as information on methods were derived from a series of morbidity surveys conducted in this country during the earlier years of this century, mostly under the auspices of the United States Public Health Service. These included studies of seven cotton-mill villages in South Carolina in 1916 (Sydenstricker *et al.,* 1918), of Hagerstown, Maryland, between 1921 and 1923 (Sydenstricker, 1926), of 9000 families selected at random from all parts of the United States between 1928 and 1931 (Collins, 1933), of over 700,000 representative households from 1935 to 1936 (Perrott *et al.,* 1939) and of the Eastern Health District of Baltimore between 1938 and 1943 (Downes and Collins, 1940).

There have been two major continuing surveys on a national basis. In England and Wales, the Social Survey, initiated in 1944, aimed to interview random samples of the population aged 16 or more (Slater, 1946). About 3000 persons were interviewed each month, and questioned as to the occurrence of illness during the previous 2 months. Routine results were published in Reports of the Ministry of Health and in the Quarterly Returns of the Registrar General, and special reports appeared from time to time

(Stocks, 1949; Registrar General, 1955). The Social Survey was discontinued in 1952.

The U. S. National Health Survey, established by Act of Congress in 1956, began collecting data in July, 1957. The program of the survey consists of three parts. The first is the household interview survey. Approximately 3000 households, selected in such a manner as to be representative of the population of the United States, are interviewed each month by trained lay interviewers (each household is interviewed only once). The sample is selected in such a way that each week data are obtained from a random sample of the population. Thus although the survey is not designed primarily to demonstrate short term time trends in disease frequency, the data may in fact be used very effectively in this way, as illustrated in Figure 7 which shows the effect of the 1957 influenza epidemic. As data accumulate from the household interview survey, more complex and more critical tabulations and cross tabulations will be possible. The fact that demographic

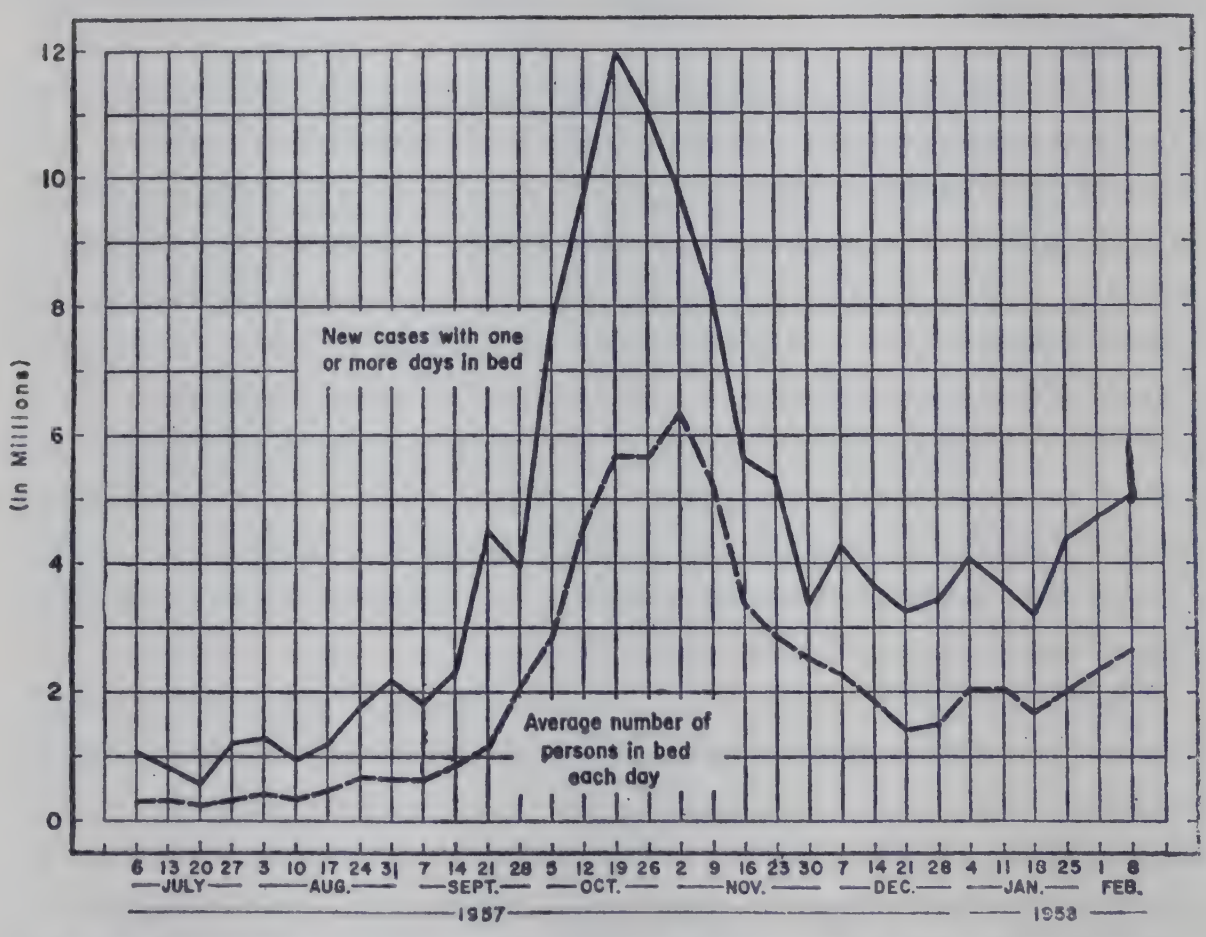

(In Millions)
0
2
4
6
8
10
12
New cases with one or more days in bed
Average number of persons in bed each day
6 13 20 27 3 10 17 24 31 7 14 21 28 5 12 19 26 2 9 16 23 30 7 14 21 28 4 11 18 25 1 8
JULY AUG. SEPT. OCT. NOV. DEC. JAN. FEB.
1957
1958

FIGURE 7. Weekly incidence and prevalence estimates of bed-disability due to acute upper respiratory disease in the United States, July 6, 1957 to February 8, 1958. National Health Survey data. (Reproduced with permission from Linder, 1958.)

and personal data as well as health information is sought at interviews ensures the derivation of the necessary "reference population" for the computation of rates in relation to variables for which population data are not presently available.

Second, a series of special studies are planned as part of the National Health Survey. Prominent among these studies are plans for a continuing survey using medical examinations as well as household interviews.

Lastly, a major part of the National Health Survey program consists of method studies designed to understand sources of error in the collection of morbidity statistics. It is perhaps not too great an exaggeration to say that the National Health Survey offers the potential for prompting as great an advance in epidemiologic knowledge in this country as did the introduction of death registration in London four hundred years ago.

7

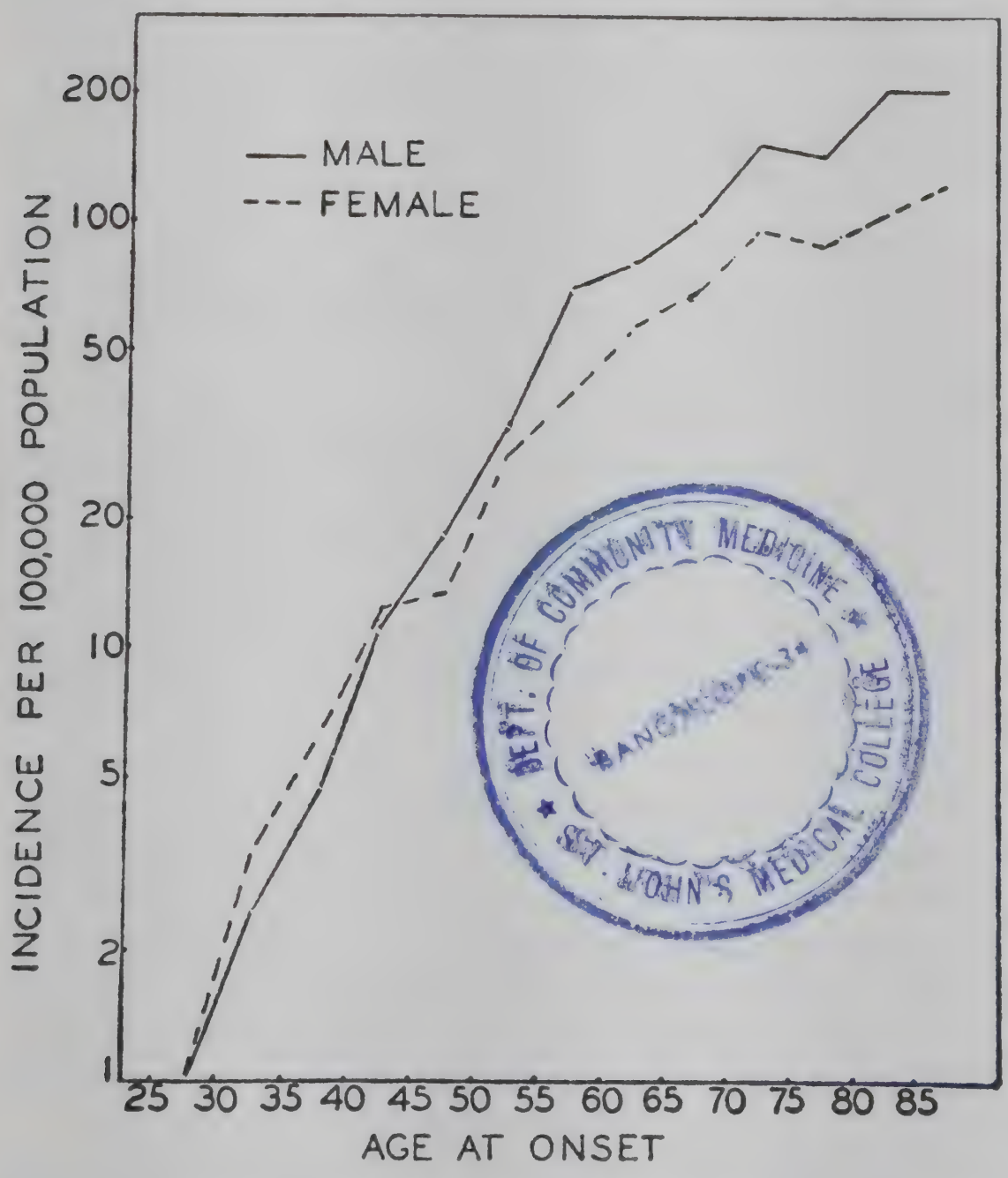
Characteristics of Persons: Age

Age is probably the most important single variable which must be considered in descriptive studies. For most diseases the variations in frequency that occur with changes in age are greater than those to be found in association with any other variable. Knowledge of age associations is therefore important for two reasons: first, study of this variation may assist understanding of the factors responsible for the development of the disease; second, associations between age and disease frequency are so strong that they produce indirect effects that must be eliminated in examination of differences in disease rates in association with other variables. Comparisons of any two populations with respect to disease frequency are generally meaningless unless account has been taken of possible age differences between them.

Association of disease frequency with age is usually measured by relating the number of cases of the disease in each age group to the population in the same age group, and deriving a succession of age-specific incidence or prevalence rates.

Frequently, and particularly in the examination of material collected from clinical sources, it is possible to present the distribution of a series of patients by age, but information is lacking as to the age distribution of the population from which the patients came. Statements such as "this disease is most common be-

tween the ages of 50 and 60" may be made on the basis of such a frequency distribution. While they have meaning to the clinician or the administrator insofar as they indicate that more of the patients seen are in this age group than in any other, such statements do not of course indicate that the greatest risk of the disease lies in this age group. Even diseases for which risk increases with age throughout life will show decreasing numbers of cases in the highest age groups because the population decreases rapidly with increasing age after about age 55. Thus the data on which Figure 8 is based showed the largest numbers of cases in



DEPT. OF COMMUNITY MEDICINE
ST. JOHN'S MEDICAL COLLEGE

FIGURE 8. Incidence of rectal cancer according to age and sex in ten metropolitan areas of the United States, 1947. (Data from Dorn and Cutler, 1959.)

the age group 55-59 for males and 60-64 for females but the highest rates are noted in the oldest age groups for both sexes. Although the necessity of using rates to express risks is frequently overlooked, the point will not be labored here in view of the discussion of rates already presented in Chapter 5.

COHORT AND CURRENT AGE CURVES

The association of a disease with age may be described in two ways:

(1) An age curve based on the age-specific rates observed in a cross section of a population at one point in time. For example, Figure 8 represents the age-specific incidence rates of rectal cancer noted in the population of ten selected areas of the United States in 1947. The populations on which the rates in the different age groups are based belong to quite different generations; the rate for the age group 30-34 is based on persons born between 1913 and 1917, that for the age group 60-64 on persons born between 1883 and 1887. This type of curve is known as a current or cross-sectional age curve.

(2) A curve based on age-specific rates for a group of persons born in a certain period and observed at successive points in calendar time as they grew older. Thus, for persons born in 1890 the incidence rate at age 30-34 would be based on data for the years 1920 to 1934, and, in the same curve, the rate for ages 60-64 would be based on data for that age group from 1950 to 1954. This type of age curve is known as a generation or cohort age curve.

The difference between these two methods of describing age associations is of importance only when the frequency of the disease under consideration has been changing over a period of time, since cohort and current age curves will be identical unless there has been secular change in rates. If, however, there has been a change in incidence and if the distinction between the two methods is not clearly kept in mind, misleading interpretations are likely. This was first illustrated by Andvord (1930), Frost (1939) and others in studies of age curves for tuberculosis mortality.

Frost examined death rates from tuberculosis in Massachusetts for the years 1880 to 1930. The data on which his analysis was

based are shown in Table 8 together with data for 1940 and 1950 added by Pope and Gordon (1955). Current age curves comparable to those of Figure 8 are derived by reading down the

TABLE 8

Death rates from all forms of tuberculosis for males in Massachusetts, 1880 to 1950†*

Age group	Year of death							
	1880	1890	1900	1910	1920	1930	1940	1950
0-4	*760*‡	578	309	209	108	41	10	4
5-9	*43*	49	31	21	24	11	2	1
10-19	126	*115*	90	63	49	21	6	1
20-29	444	361	*288*	207	149	81	36	7
30-39	378	368	296	*253*	164	115	55	25
40-49	364	336	253	253	*175*	118	86	42
50-59	366	325	267	252	171	*127*	95	69
60-69	475	346	304	246	172	95	*108*	87

* Rates are annual death rates per 100,000 male population.
† Data from Frost (1939) and Pope and Gordon (1955).
‡ Rates for the cohort of 1875 are italicized.

columns; they are drawn in Figure 9. Frost pointed out that in current age curves the peak of mortality shifted to a higher age in successive decades. Cohort age curves may be drawn from the same data by reading diagonally across the table from top left to bottom right; they are shown in Figure 10.* All the curves now have a consistent pattern. They indicate that for any given generation, after the initially high rates in the 0-4 year age group, the highest tuberculosis death rates occurred in the 20-29 year age group.

Two interpretations of these findings are possible. The first is that the marked fall in the mortality from tuberculosis was in fact accompanied by a change in the association of the disease with age — that the factors responsible for the decline in mortality acted differentially at different ages. The second is that there was no change in the age association of the disease during the period of decline — that the decline affected all age groups equally —

* Note that the exact data for cohorts are not contained in the table. Thus, for the age group 0-4 for the cohort of 1875 data for 1877 to 1878 would be desirable, since this cohort occupied the center of the age group in these years. The data for 1880 are used as an approximation.

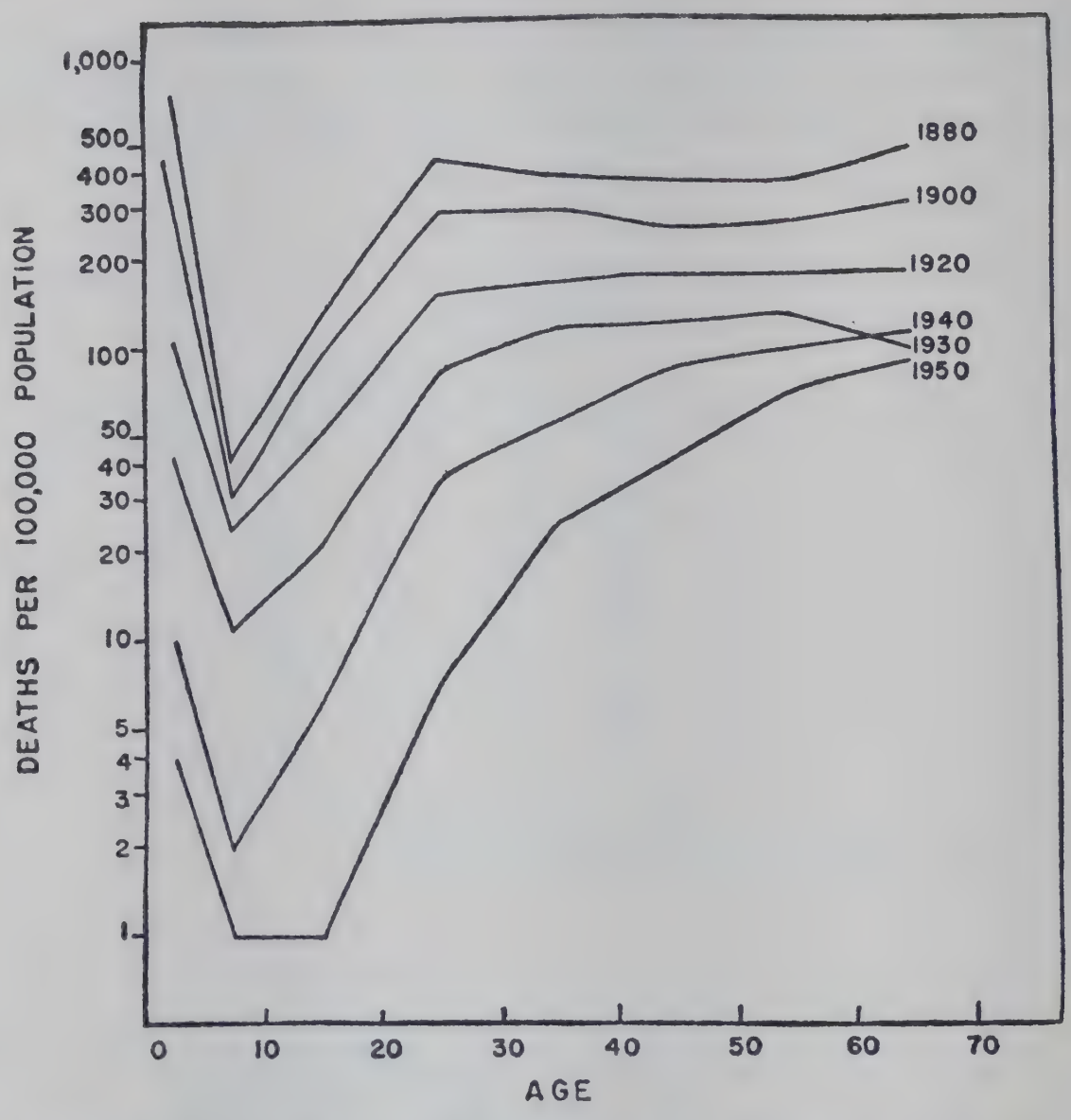


FIGURE 9. Tuberculosis mortality according to age (current age curves), Massachusetts males, 1880 to 1950. (Data from Table 8.)

but that the mortality level for any age group was characteristic of the cohort to which the age group belonged rather than of the period during which the mortality was measured. (Note that *age association* here means an association of disease mortality with age which is independent of mortality changes which have occurred over time.) The latter interpretation was accepted by Frost and others who considered the question. It involves the supposition that the mortality level of a cohort is determined very early in its existence. This supposition is consistent with the hypothesis that the mortality from tuberculosis in adult life is predominantly endogenous — resulting from the activation of latent foci — and determined for any cohort largely by the degree of infection received in early life.

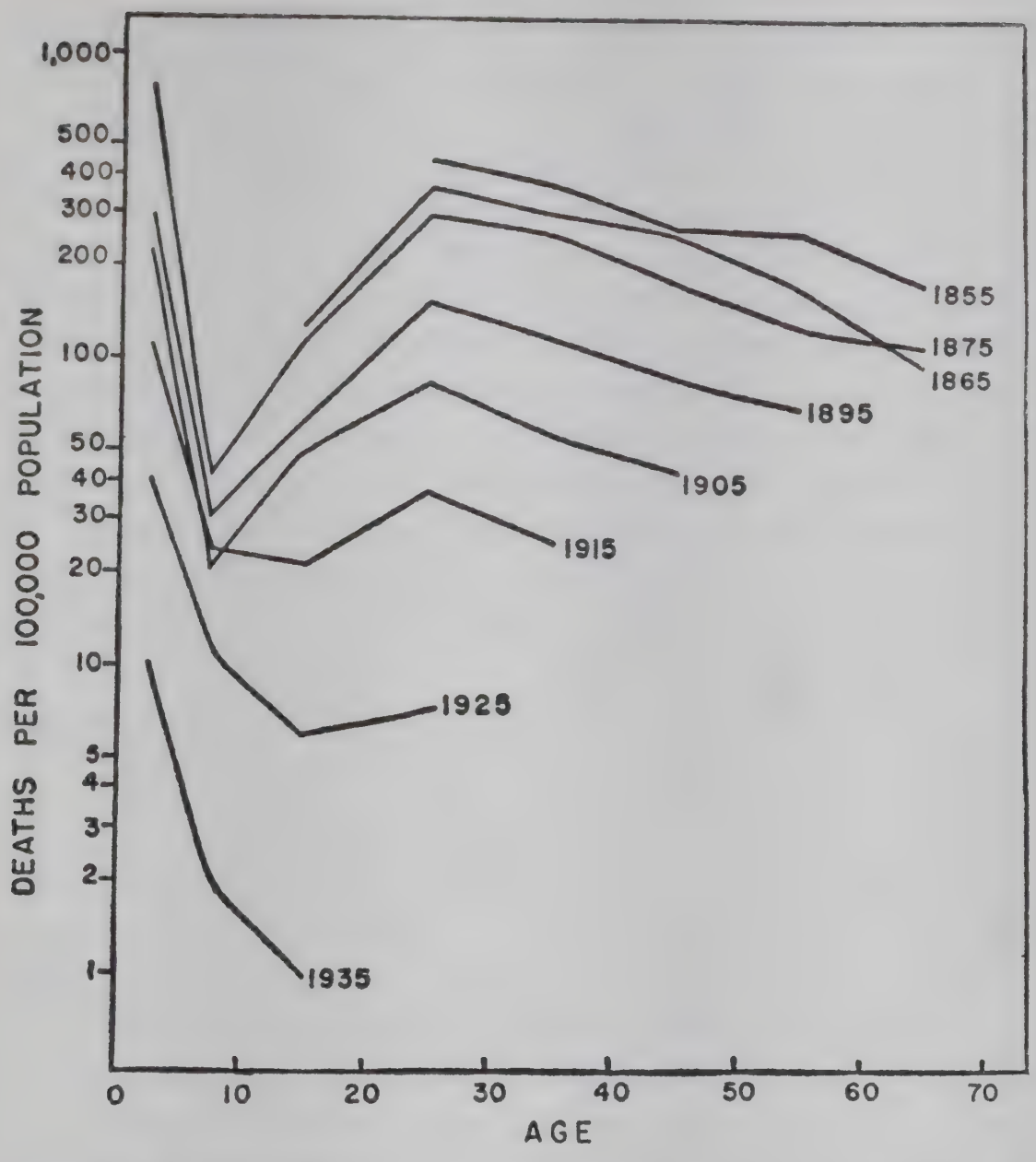


FIGURE 10. Data from Table 8 arranged to show the mortality experience of Massachusetts male cohorts born between 1855 and 1935 (cohort age curves).

This is an example of a situation in which current age curves were judged to have been distorted by rate changes which had taken place during the past. Current age curves appeared distorted because the age association seemed best described by a cohort curve. However, cohort curves may be similarly distorted if current curves are more appropriate for the description of a particular situation.

In some situations, a decision may be made on a priori grounds as to the type of analysis — cohort or current — which will most appropriately describe the age association of a particular disease. Thus, an epidemic of certain acute infectious diseases will affect

all or most cohorts living at the time of the epidemic regardless of their dates of birth. For example, in a study of the mortality rates of successive cohorts in England and Wales, Case (1956) noted the effect of the 1918-1919 influenza epidemic as an upward kink in the age curve at ages 15-19 for the cohort born in 1901, at 25-29 for the cohort born in 1891, and at 35-39 for the cohort born in 1881. The fact that different cohorts were affected at different ages but that all were affected during the same calendar time suggested, as expected, that the change was characteristic of a time and not of a cohort. Note that here any single cohort age curve is misleading, since it suggests an "age effect" at the age at which that cohort was affected by the epidemic. Thus, the kink in the age curve at ages 15-19 for the cohort of 1901 had in reality nothing to do with age but was merely a circumstance of the cohort being of that age at the time of the epidemic. Misinterpretation of current age curves will, in theory, not be possible under such a circumstance, since risks for all age groups will be based on years of equal epidemicity.

Suppose, however, that an epidemic of a disease occurred in early life and that it was associated with chronic or permanent residua, as, for example, maternal rubella followed by congenital cataract. In such a situation high prevalence rates for the anomaly would be found at all ages of the affected cohort, but (if the epidemic were not repeated) there would be no period at which more than one age group was affected. In this case, the cohort age trend would best describe the age association, since for any cohort all values will be based on populations experiencing equal degrees of exposure, while current curves will indicate (incorrectly) an age effect in the age group occupied by the affected cohort at the calendar time referred to by the particular current age curve.

In the examples of influenza and rubella the propriety of applying the one or the other type of analysis was apparent. Commonly, however, the issue is not clear-cut. For example, in tuberculosis the hypothesis that attack rates experienced by a cohort early in life determine its level of tuberculosis mortality throughout its existence seems neither more nor less reasonable than the assumption that tuberculosis mortality depends on environmental conditions existent closer to the time at which the mortality

is being measured. Under the first hypothesis, cohort curves will reflect the age association more accurately, but under the second, current curves would be indicated. While an a priori decision between the two methods cannot be made in this case, the remarkable consistency of the age curves derived by the cohort method suggests that the changes are in fact occurring in a pattern consistent with the first hypothesis. To be consistent with the second, a subsidiary hypothesis needs to be introduced to the effect that age groups have been unequally affected by changing environmental conditions. As noted by Springett (1952), evidence as to the applicability of one or the other type of analysis will depend on both (1) a priori reasoning, based on knowledge of the pathogenesis and natural history of the disease in question, and (2) demonstration of a more consistent age pattern in one or the other type of analysis.

It is sometimes assumed, incorrectly, that one or the other type of curve (usually the cohort curve) is superior for all situations. For example, Korteweg (1951) and others have noted that age incidence and mortality curves for lung cancer as derived from current data show a peak at ages 60-69 and thereafter decrease with age, while cohort age curves do not show this fall, the risk increasing throughout the life span. Illustrative data are shown in Figure 11. In commenting on these findings, several authors have suggested that the cohort curves portray the real change in risk with age, and that the pattern revealed by the current curve results from the striking increase in incidence of lung cancer which has occurred during this century. However, this assumption may not be justified. For example, a person age 70 born in the cohort of 1880 achieved that age in 1950; a person age 60 of the same cohort achieved that age in 1940. Either an increase between 1940 and 1950 in the risk of acquiring lung cancer or an increase in the chances of the disease being diagnosed could account for an increase in the rates between age 60 and 70. Such an increase would be a function not of age but of time, and the cohort age curve would therefore be misleading if interpreted solely as a function of age. On a priori grounds, the hypothesis that lung cancer rates are in part characteristic of a cohort is certainly a valid one in view of the large body of experimental evidence on the long latent period known to be associated with car-

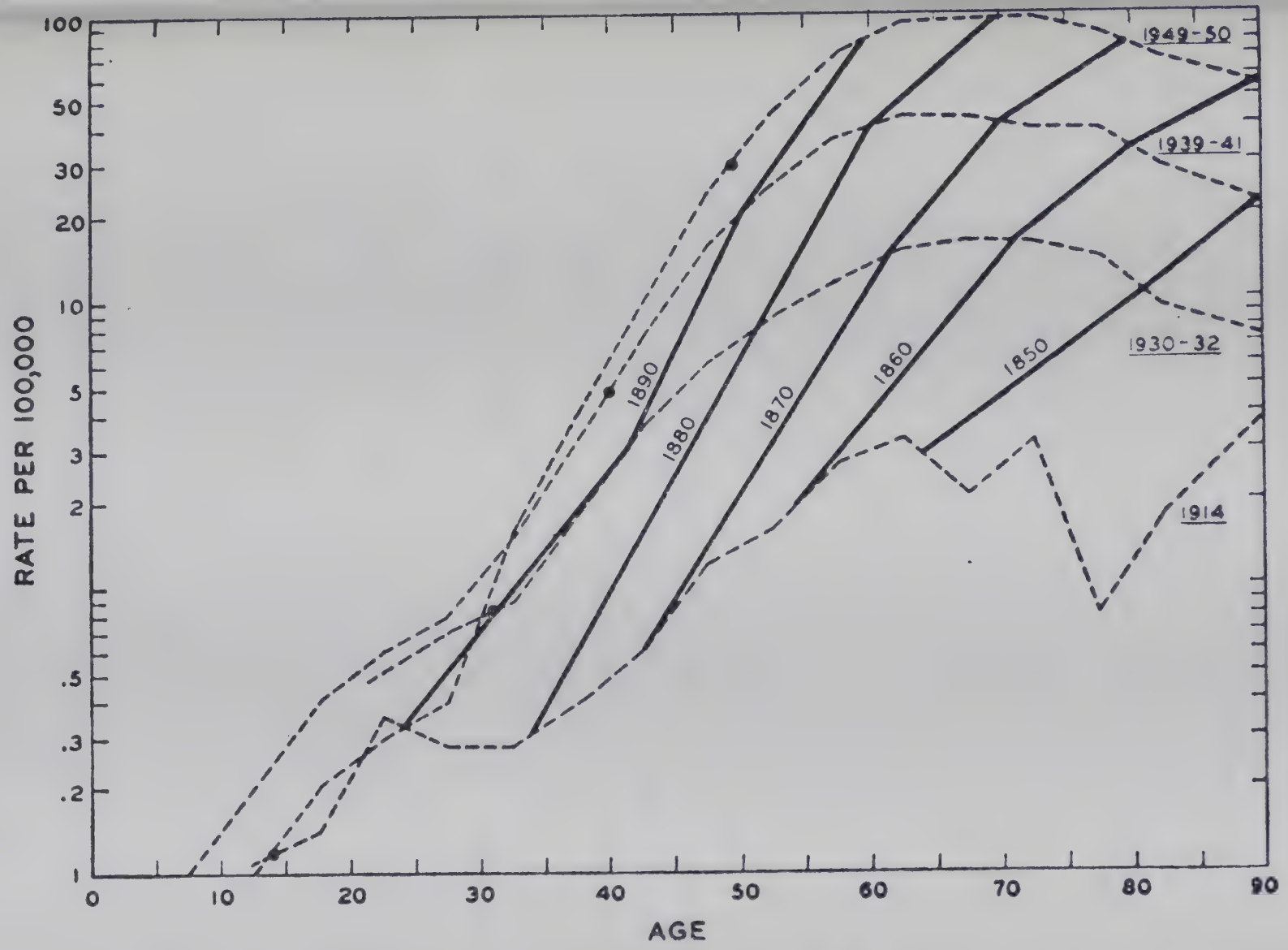


FIGURE 11. Mortality rates for cancer of the lung and bronchus, white males, United States, 1914 to 1950. (Reproduced, with permission, from Dorn and Cutler, 1959.) (Solid lines are cohort curves, broken lines are current age curves.)

cinogenesis, but there are also features of cancer morbidity, such as completeness of diagnosis, which might be expected to be characteristic of a certain period rather than of specific cohorts. Equally valid grounds for the use of either current or cohort curves can therefore be advanced. In addition, neither type of curve can be said to reveal more "consistent" results than the other. The age curve for lung cancer shows no convenient peaks, such as seen in the tuberculosis cohort curves, which would assist in the decision. Neither can the method of approaching the problem discussed in connection with changes in disease rates that occur over time (see Chapter 11) be applied to the data on lung cancer, for reasons discussed in that chapter (page 176). There is therefore, in this particular instance, no way of deciding which of the two types of curves illustrated in Figure 11 is the more accurate indicator of the age association of lung cancer.

Implications from Cohort Analysis

If a decision can be made, whether on a priori grounds, by the demonstration of a more consistent age pattern, or by the method discussed in Chapter 11, that incidence changes are apparently occurring in a cohort pattern, certain implications follow.

In the first place, apparently paradoxical trends such as those noted for tuberculosis in Figure 9 are explained. In diseases, such as the neoplastic diseases, in which incidence increases throughout the life span the distortion of the age trend may not be so obvious as in diseases, such as tuberculosis, whose peak incidence occurs during early adult life. In such cases the distortion may amount merely to a steepening or diminishing of the slope of the age trend. The theoretical situation is illustrated in Figure 12. If incidence over time is increasing, and the increase is occurring in such a way as to affect successive cohorts, then a current age study tends to reduce the slope of the curve showing the actual association with age. If incidence is decreasing in a similar fashion, then a current age study exaggerates the slope of the curve of age association. Conversely, when the age association is characteristic of a period of time rather than of cohorts of the population, then a cohort study exaggerates the association if incidence is increasing and minimizes it if incidence is decreasing. An example of one of these situations is evident in Figure 11. Cohort

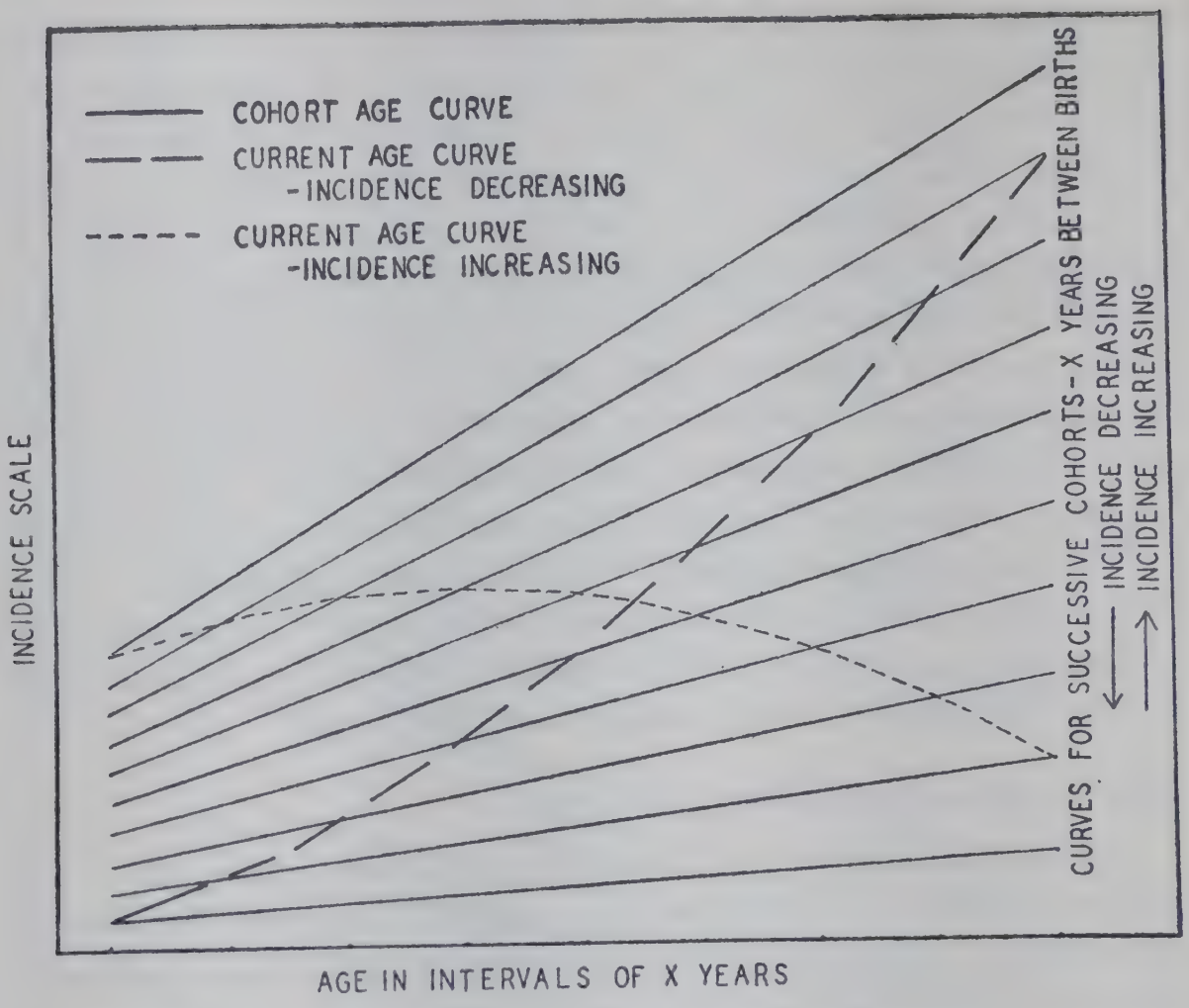


FIGURE 12. Schematic comparison of cohort and current age curves under conditions of increasing and decreasing incidence.

age curves for lung cancer show a steeper slope than do current curves because of the increases in incidence which have occurred over time.

The fact that age curves for a disease follow a cohort pattern influences hypotheses concerning the factors responsible for changes in disease incidence over time. Thus, as noted by Frost with respect to tuberculosis, it implies that the factors are operative early in the existence of the cohort. In this case the supposition is that the decreasing infection rates in the first 5 years of life resulted in decreased rates throughout the remainder of life. It is pointed out in Chapter 11 that time changes in several neoplastic diseases have also followed a cohort pattern. In these latter diseases the evidence suggests that the environmental situations or personal habits which have been responsible for changes in the incidence of these neoplasms become established prior to age 35 — the age at which the first differences between

cohorts are evident. It should be noted that this implication applies only to the *changes* in disease incidence which have occurred over time. Thus, conformity to a cohort pattern does not imply that *all* the factors responsible for later tuberculosis mortality are operative in the early years. There undoubtedly are factors operative later in life also. The implication is, however, that the latter factors have not changed to any considerable extent over the period examined, and that the *change* in incidence at later ages has resulted primarily from changes in factors operative early in life.

INTERPRETATION OF AGE ASSOCIATIONS

Innumerable factors enter into the determination of the shape of the age association curves of a given disease. Only a few of these are understood.

At the most superficial level, there are a number of circumstances which in fact have little to do with age per se, but which may nevertheless affect the shape of an age incidence curve. One of these — the use of current age curves when cohort curves are appropriate, and vice versa — has already been discussed. Variations in accuracy of diagnosis according to age must also be considered in this category, particularly in regard to the oldest age groups. There is, in many clinical situations, less concern over establishing an exact cause of death for an 80-year-old than for a 40-year-old person. This situation tends to reduce age-specific rates for the oldest age groups for diseases of difficult diagnosis, and to exaggerate rates in the same age groups for "wastebasket" diagnoses. In addition, basic population data may be inaccurate for the oldest age groups, leading to further distortions. The latter situation applies in the United States particularly to the non-white population.

If the rate of progression of a disease varies with age, age-risk curves may vary according to whether they express rates of onset, diagnosis or mortality. For example, in a disease in which the average interval between onset and death increases with age, a mortality age curve will in general be less steep than an incidence curve. Even when case fatality and average survival time are the same for all ages, the mortality curve (shifted, of course, to-

wards the right of the age scale) may be different in shape from the incidence curve for the same disease. Thus, if a final inclusive age group is used (e.g., 75 and over), mortality rates will be higher than incidence rates in that age group while the reverse will be true in the earlier age groups. An illustration of the differences between age curves based on incidence, prevalence and mortality data is shown for a chronic disease — multiple sclerosis — in Figure 13.

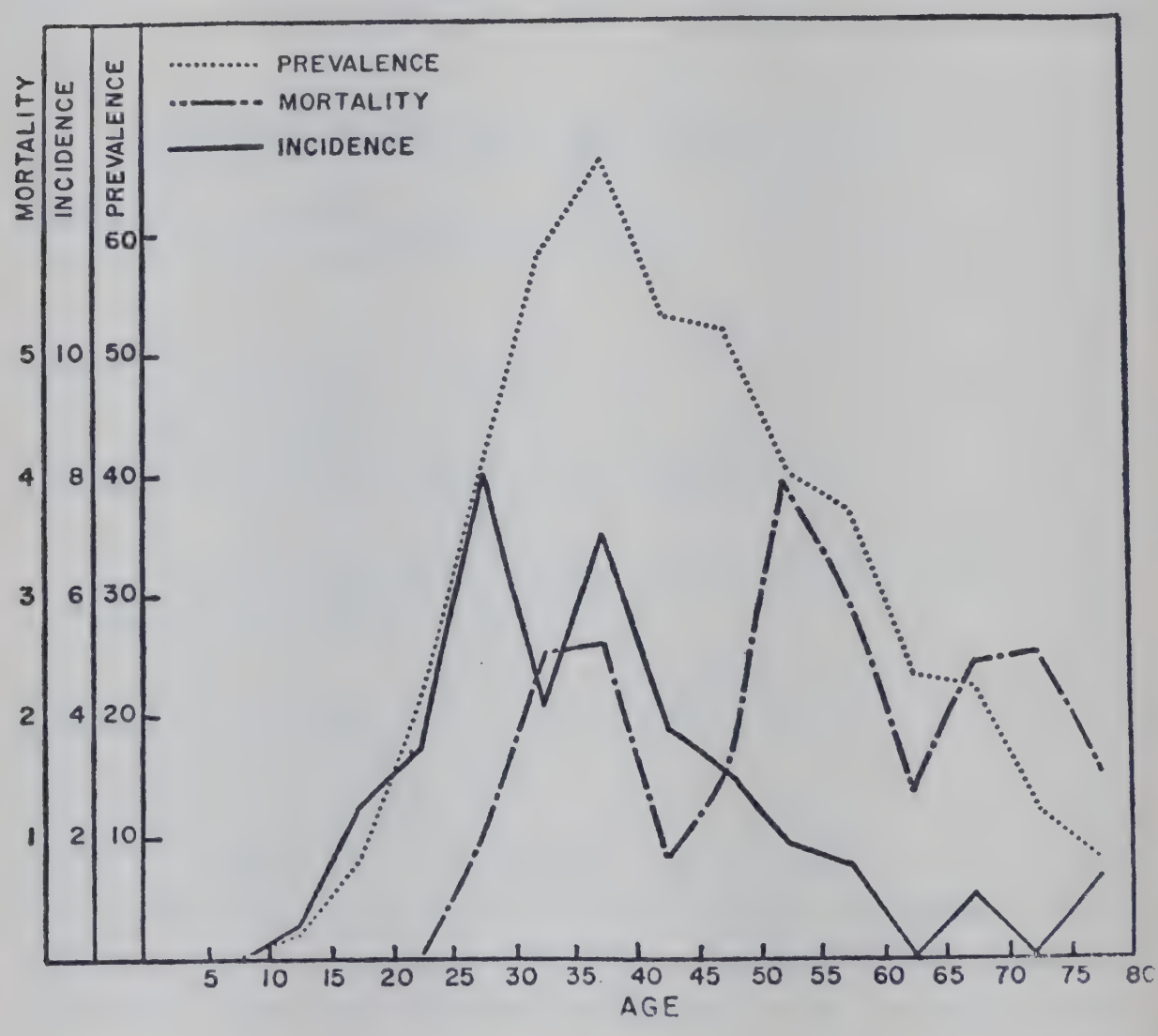


FIGURE 13. Rates of incidence, prevalence and mortality per 100,000 of multiple sclerosis in males, Boston, 1939-1948. (Data from Ipsen, 1950.)

If the disease under examination is of such a very chronic nature that a considerable proportion of those affected die from other causes before death occurs from this disease, then the mortality age curve of the disease will be influenced by the current age-specific risks of dying from all causes — provided, of course,

that all cases in which the disease is present at the time of death are used in the calculation of the curve. This is merely a restatement that mortality data are subject to misinterpretation if the disease is not one in which death occurs in a high proportion of the patients, and within a reasonably short time of onset.

Bimodality — that is, the occurrence of two separate peaks — in an age incidence curve of a disease is always of interest. It indicates, first of all, that the material is not homogeneous — that the entity under examination might properly be divided into two. Even when an experiential, rather than a manifestational, entity is being examined, bimodality suggests the existence of experiential differences other than that on which classification of the disease is based. For example, the age curves for tuberculosis in Figure 10 (page 91) are bimodal, showing one mode in the 0-4 age group and a second in the 20-29 age group. Since tuberculosis is defined in terms of experience with the tubercle bacillus, these two modes are both affecting the same "disease entity." This epidemiologic finding suggests that two distinct sets of ancillary experiences must be taken into account to explain the entity's observed age distribution.

Age incidence curves for the manifestational entity Hodgkin's disease are also distinctly bimodal, showing one peak in early adult life and a second in old age (Figure 14). Probably, therefore, parts of two distinct experiential entities are congruent with this manifestational entity as presently constituted. This hypothesis has been supported by the demonstration that the disease in persons under 40 years of age differs from that in persons over that age with respect to sex ratio, racial, religious, and geographic distribution, and duration of illness (MacMahon, 1957). Apart from suggesting the presence of separate experiential entities, such evidence can be of service in the finer definition of manifestational entities if clinical and pathologic findings are re-examined in the light of the epidemiologic differentiation.

A common type of age curve is that exhibited by the so-called degenerative diseases in which incidence rates increase progressively throughout the life span. Understanding of the mechanisms responsible for this type of age curve is a problem which is only in part epidemiologic. At present it seems likely that fundamental studies of an experimental type have most to offer in the

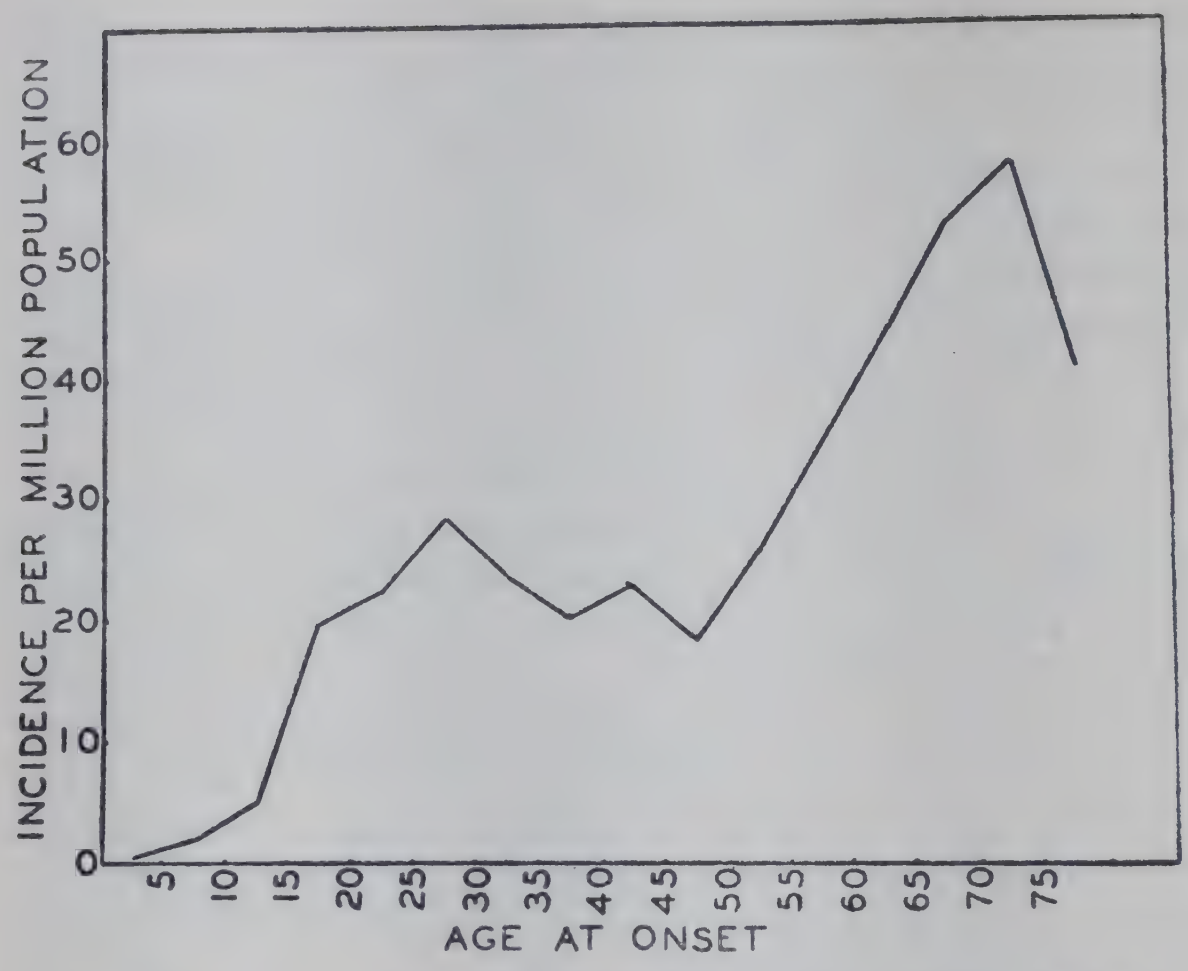


FIGURE 14. Incidence of Hodgkin's disease in the white population of Brooklyn, 1943 to 1952. (From MacMahon, 1957.)

understanding of those bodily changes which occur with age and which lead to increasing rates for so many different diseases. Probably the explanations of such associations are as numerous as the diseases which exhibit them. Certainly change in no single factor is likely to be responsible for incidence changes in so many fundamentally different disorders.

In considering the age curves for neoplastic disease, Huxley (1951) has suggested that they result from a genetic selection which reduces the number of persons susceptible to tumors genetically destined to occur prior to the end of the reproductive years. This is clearly not a complete explanation, since those people or tumors surviving such an evolutionary process must possess some characteristic which destines onset of the tumor after the reproductive age. We are still faced with the elucidation of this characteristic. There is little evidence to support Huxley's hypothesis; the original premise on which it was based — that tumors which occur during the reproductive years are less common than

those occurring later — has been shown to be untrue (Landauer and Beall, 1952).

It is not known whether there are truly endogenous changes with age which affect susceptibility to disease — that is, changes which are the inevitable consequences of the passage of time and which are completely independent of environmental circumstances to which the individual is exposed during that time. Again with reference to neoplastic disease, Hueper (1952) has supported the view that the shape of the age curve is determined largely by the length of the latent period between encounter with carcinogenic agents and manifest neoplasia, by the age at which such exposure first takes place, and by the intensity of exposure. The long latent periods found in human neoplastic disorders following exposure to known carcinogenic agents (up to 50 years) make it inevitable that such neoplasms should occur in the older ages, even if exposure was first effective at birth. Hueper points out that the age distribution of scrotal cancer and other occupational neoplasms varies according to the age at first exposure to the carcinogenic agent. In addition, there is no experimental evidence that the latent period between initiation of exposure and development of neoplasia is any longer for young than for old experimental animals. There is therefore a good deal of support for Hueper's hypothesis.

However, this view is not so different as it might appear to be at first sight from the hypothesis of changing tissue susceptibility with age. "Aging," as a characteristic of any individual, is merely the passage of time; this passage of time must necessarily be accompanied by exposure to environmental situations which alter the body chemistry and physiology. Among the most obvious of these situations are those necessary for the continuation of life such as the intake of food and oxygen. Some of these alterations may result in increased incidence of certain diseases. If the changes result from exposure to a known toxic agent, and if such exposure is avoidable and therefore not a necessary part of the aging process, such as, for example, in the case of occupational carcinogens, we would not regard this as being a primary association with age, even though aging is essential to the accumulation of a sufficient dose of the toxin. If, however, the disease-inducing changes were produced by environmental situations which

MP 100
N60
05410
COMMUNITY HEALTH
LIBRARY
AND

were essential concomitants of the aging process, such as, for example, prolonged consumption of oxygen, should this be regarded as a primary effect of age or not? Certainly, aging would inevitably be accompanied by such changes. Between these extremes are a wide variety of situations in which environmental circumstances may be separated to a greater or lesser degree from aging. For example, it may be that some component of the incidence of human leukemia can be attributed to exposure to background radiation from natural sources. At present aging is necessarily accompanied by cumulative exposure to background radiation, and the relationship between age and leukemia incidence may, under this hypothesis, be regarded as a fairly direct one. However, if methods of protection were devised such that only artificial sources of radiation contributed to the total radiation experience, leukemia incidence as influenced by background radiation might be regarded as independent of age.

Whether or not there is a truly endogenous aging process (a "biological clock") resulting in increased disease susceptibility, the evidence assembled by Hueper indicates that there are certainly many alterable situations, separable from age, which contribute to the shape of age incidence curves. The existence of a progressive association with age cannot *ipso facto* be regarded as evidence of endogenous causation and thus as grounds for pessimism with regard to preventive possibilities. Practically, effort should be concentrated on the elucidation of those deleterious environmental exposures which accompany but are nevertheless separable from age. When these have been eliminated, greater concern may be felt over the possibility of an endogenous component in the aging process.

8

Characteristics of Persons: Sex and Ethnic Group

This chapter continues the discussion of variables that are descriptive of the types of person affected by disease. It is concerned with two variables, sex and ethnic group, which in conjunction with age are commonly thought of as the basis of descriptive epidemiology. Their examination here is intended to illustrate the types of epidemiologic analysis and interpretation that can be applied to descriptive variables in general.

SEX

Measurement of Association

As with descriptive variables other than sex, the association between a particular disease and the variable is most satisfactorily demonstrated by a difference in disease rates between subgroups of the variable — in the case of sex between disease rates for males and for females. In an easier (and commonly used) procedure deviations in the sex ratio* from 50 per cent are noted, under the assumption that the sex ratio in the population from which the cases were drawn is also 50 per cent. It is a fact of life, however its justice may be argued, that this assumption may in

* Number of affected males as a percentage of the total affected.

specific instances not be entirely correct. For example, among the 1955 Massachusetts population 70 or more years of age the sex ratio was only 44 per cent. Calculation and comparison of sex-specific rates are therefore preferable in that one fewer assumption needs to be considered in the analysis of findings.

In most populations the two sexes also have different age distributions, females commonly being on the average older than males. Small differences in sex-specific rates must therefore be viewed with caution unless the possible influence of age differences has been taken into account.

Sex Distribution of Disease

More males are born alive than females in a ratio of about 106:100. Death rates have consistently been higher for males than for females, except at child-bearing ages in situations where standards of obstetric care have been poor, when rates for females have been observed to exceed those for males. The disparity between male and female death rates in the United States at various ages is shown for 1955 in Table 9. Of interest is the fact that the death rates for United States males are currently highest relative to those for females during the child-bearing ages of females.

TABLE 9

Sex-specific mortality rates in broad age groups in the white population of the United States, 1955*

Age group (years)	Mortality rates			Male rate as percentage of female rate
	Total	Male	Female	
Under 1	26.2	29.8	22.4	133
1-4	1.0	1.1	0.9	122
5-14	0.5	0.6	0.4	150
15-39	1.4	1.8	0.9	200
40-64	9.1	12.0	6.2	194
65 and over	61.5	71.7	52.8	136
All ages†	8.9	10.9	7.2	151

* Rates are annual death rates per 1000 population.

† For "all ages" and within each broad age group the rates are standardized by the direct method to the age distribution in 5-year age groups of the total United States population in 1955.

In the United States the general disparity has been increasing. Illustrative data for 1933* and 1955 are presented in Table 10. Comparison of age-standardized mortality rates in the two years indicates that the increased difference results because the rates for females fell more rapidly than those for males.

TABLE 10

Mortality rates for males and females of the white population of the United States, 1933 and 1955*

Year	Mortality rates		Male rate as percentage of female rate
	Male	Female	
1933	14.0	11.5	122
1955	10.9	7.2	151

* Rates are annual death rates per 1000 population, and are age-standardized as in Table 9.

The less fortunate male position is not the result of serious disadvantage in only a few of the assigned causes of death (Table 11). Although considerable variation in sex pattern is evident between assigned causes, death rates among males exceeded those among females for the majority of the causes, commonly by as much as 150 per cent or more, and in one instance (respiratory cancer) by 585 per cent. Among the commonest causes of death, excluding deaths assigned specifically to diseases of the sex organs, only diabetes mellitus and hypertensive heart disease showed higher death rates among females than among males.

A seemingly paradoxical finding in surveys of morbidity (rather than death) reveals that illness rates, defined in a variety of ways, have consistently been found higher among females than males. Data from four such surveys are shown in Table 12. In spite of gross variation in definitions of illness and incapacity, and the different measures of frequency used, rates for females are consistently and substantially higher than those for males. This consistency indicates either that females admit to illness more readily than males or that they experience a higher incidence of illnesses which, in view of the preceding mortality data, must presum-

* This year is chosen for comparison because it is the first year in which data for the whole United States are available.

TABLE 11

Mortality rates from selected common causes for males and females in the United States, 1955*

Assigned cause	Mortality rates			Male rate as percentage of female rate
	Total	Male	Female	
Respiratory cancer	18.0	31.6	5.4	585
Peptic ulcer	5.9	9.6	2.5	384
Suicide	10.1	16.2	4.5	360
Homicide	4.5	7.0	2.1	333
Motor vehicle accidents	23.3	35.9	11.4	315
Syphilis	2.3	3.5	1.2	292
Tuberculosis	9.1	13.3	5.2	256
Urinary cancer	7.0	9.8	4.6	213
Cirrhosis of liver	10.2	13.8	6.7	206
Accidents, other than motor vehicle	34.5	46.1	23.2	199
Arteriosclerotic heart disease	251.6	336.0	174.4	193
Appendicitis	1.4	1.8	1.0	180
Digestive cancer	52.4	61.2	44.6	137
Vascular lesions of the central nervous system	109.8	115.1	105.3	109
Hypertensive heart disease	46.1	44.5	47.2	94
Diabetes mellitus	15.5	13.1	17.7	74
Genital cancer	23.0	19.7	26.8	74

* Rates are annual mortality rates per 100,000 population. They are age-standardized as in Table 9.

ably be less lethal than those experienced by males. Data from the U. S. National Health Survey (1959) show that this sex pattern for morbidity is evident in all gradations of socio-economic status, in urban and rural areas, and when the examination is restricted to gainfully employed persons. When separate causes of illness are considered, women have higher morbidity rates than men for the majority of chronic diseases. Many chronic conditions are strikingly more common in women — for example, thyrotoxicosis, diabetes mellitus, cholecystitis and biliary calculi, obesity, arthritis, psychoneurosis and many benign tumors. Only a few illnesses appear to be predominantly masculine diseases when examined in morbidity data; these include peptic ulcer, inguinal hernia, accidents, arteriosclerotic heart disease and respiratory cancer.

TABLE 12

Morbidity rates for males and females in four surveys

Survey	Measure of illness or incapacity	Rates per 1000	
		Males	Females
U.S. "9000 families" survey, 1928-1931 (Collins, 1940)	Total illnesses (annual prevalence)	720	915
	Nondisabling illnesses (annual prevalence)	285	374
	Disabling illnesses (annual prevalence)	435	541
	Bed illnesses (annual prevalence)	345	476
	Days sick per year	23217	35241
	Days in bed per year	3046	4699
Baltimore "Eastern Health District" survey, 1938-1943 (Downes, 1950)	Acute illness (annual prevalence)	1044	1474
	Chronic illness (annual prevalence)	191	285
	Doctor's calls (annual incidence)	2123	2706
	Hospital admissions (annual incidence)	67	83
British "Survey of Sickness," 1946-1947 (Stocks, 1949)	Total illnesses (monthly prevalence)	786	1094
	Total sick persons (monthly prevalence)	487	588
	New major illnesses (monthly incidence)	46	52
	New minor illnesses (monthly incidence)	259	328
	Days incapacity in month	1079	979
	Medical consultations in month	399	420
U.S. National Health Survey 1957-1958 (1959)	Acute illness (annual incidence)	2475	2720
	Restricted-activity days (annual)	17700	22200
	Bed-disability days (annual)	6900	8700
	Hospital discharges (annual incidence)	74	83

ETHNIC GROUP

A variety of terms might be selected to distinguish groups of persons who have a greater degree of homogeneity than the population at large, whether from some degree of homogeneity of present-day customs or from a common biologic inheritance. While the use of the word "race" extends from a "permanent variety of the genus Homo" on the one hand to groups with as little common background as "the race of doctors" (Webster) on the other, "race" is so often equated with the former rather than the latter usage that we prefer "ethnic group" as conveying more clearly the intended epidemiologic concept.

Actually, the epidemiologist is interested in ascertaining disease rates among any groups whose members have lived together and/or intermarried for a sufficient length of time to have acquired common characteristics whether by biologic or by social

mechanisms. For practical reasons such groups must be identified by commonly applied descriptive terms. These include race (in its restricted sense), nativity, religion and local reproductive and social units, such as tribes and geographically or socially isolated groups.

In the discussion of sex differences, mention was made of the uncertainty of inferences derived from sex ratios rather than from sex-specific rates. A similar situation arises frequently in the comparison of ethnic groups when either of two methods is used:

(1) The frequency of persons of a given ethnic group with a specified disease is expressed as a percentage of the number of the same ethnic group with all diseases or a certain group of diseases, and these frequencies are compared between ethnic groups.

(2) The frequency of persons of a certain ethnic group with a certain disease is expressed as a percentage of the total number of persons with the same disease and compared with similar percentages among patients with other diseases or among all patients seen at a given hospital.

TABLE 13

Mortality rates from cancer in Bantu mine workers and American Negroes†*

Site	Bantu	American Negro
Liver cancer‡	12.7	3.0
Other cancer	1.3	61.5
All cancer	14.0	64.5

* Rates are annual mortality rates per 100,000 population, and are not age-standardized.

† From Gilliam, 1954.

‡ This includes cancer of the biliary passages.

Both these methods are likely to result in incorrect inferences, since the population at risk is ignored. For example, Berman (1940) noted that 91 per cent of cancers among Bantu laborers in Witwatersrand gold mines were primary cancers of the liver. Since the liver contributes only about 1 per cent of cancers among European whites, it was inferred that the Bantus suffered an extraordinarily high rate of primary liver cancer. However, by the use of race-specific rates, Gilliam (1954) has shown that the extraordinary feature of the Bantu experience is not the high fre-

quency of liver cancer but the striking rarity of all other forms of cancer. While the Bantus studied do have somewhat higher rates of liver cancer than American Negroes (and European whites), the excess is not nearly as great as that suggested by the comparison of relative frequencies (Table 13).

Race

A widely used concept of the word "race" is in its application to the major subdivisions of man as they exist at the present time — White, Negroid and Mongoloid. Much epidemiologic work has been done on the comparison of disease rates in "racial" groups defined in this way. In particular, in the United States, differences in disease rates between the white and Negro populations have been extensively studied and have been productive of hypotheses. Mortality statistics contain a great deal of information on this point, some of which is shown in Tables 14, 15 and 16. As in the case of the information shown in these tables, most of the published mortality data are tabulated to show only rates for "whites" and "non-whites," without further subdivision of the non-white population. However, since approximately 95 per cent of the non-white population of the United States is Negro, rates for "non-whites" can be considered to reflect the rates for Negroes with some accuracy.

In Table 14 it is seen that the total age-standardized mortality

TABLE 14

Mortality rates for whites and non-whites in broad age groups in the United States, 1955*

Age group (years)	Mortality rates			Rate in whites as percentage of rate in non-whites
	Total	White	Non-white	
Under 1	29.6	26.2	51.4	51
1-4	1.1	1.0	2.0	50
5-14	0.5	0.5	0.6	83
15-39	1.6	1.4	3.2	44
40-64	9.8	9.1	17.2	53
65 and over	61.5	61.5	61.6	100
All ages	9.3	8.9	12.5	71

* Rates are annual mortality rates per 1000 population and are standardized as in Table 9.

rate for non-whites was some 35 per cent higher than that for whites in 1955. Table 15 indicates that mortality rates have fallen considerably between 1933 and 1955 in both groups, the fall being somewhat greater for non-whites than for whites.

TABLE 15

Mortality rates for whites and non-whites in the United States, 1933 and 1955*

Year	Mortality rates		Rate in whites as percentage of rate in non-whites
	White	Non-white	
1933	12.7	20.4	62
1955	8.9	12.5	71

* Rates are annual mortality rates per 1000 population, age-standardized as in Table 9.

Table 16 shows rates of death assigned to selected causes in whites and non-whites in 1955. The over-all death rates in the two groups are derived from widely divergent patterns. Among the patterns that have stimulated the greatest epidemiologic interest are the high death rates from poliomyelitis in whites contrasted with the higher rates of several other infectious diseases (tuberculosis, syphilis, whooping cough) in non-whites, the higher rates of hypertensive heart disease and vascular lesions of the central nervous system in non-whites contrasted with the higher rate of arteriosclerotic heart disease in whites, and the differences observed for cancers of various sites. In addition to the differences in the common sites of cancer shown in Table 16, cancer rates for non-whites are unusual compared to those of whites in showing high rates of penile cancer and low rates of testicular and skin cancer. A few diseases show virtual restriction to one or the other racial group. The best-known example, and perhaps the commonest, is sickle-cell anemia which occurs only in persons of Negro ancestry.

As previously noted (page 105), sex patterns in disease occurrence are different in morbidity data from those suggested by mortality data. Unfortunately, there have been very few data available on racial differences in morbidity patterns, several of the important morbidity studies having been restricted to one or the

other racial group in the interests of obtaining a homogeneous study population (Collins, 1933; Downes and Collins, 1940). Also, data on racial differences have not yet been published from the U. S. National Health Survey, although the sampling procedure does not exclude any racial or ethnic group; no doubt this will eventually be a valuable source for such information.

TABLE 16

Mortality rates from selected causes for whites and non-whites in the United States, 1955*

Cause	Mortality rates		Rate in whites as percentage of rate in non-whites
	White	Non-white	
Poliomyelitis	0.7	0.2	350
Suicide	10.7	4.2	255
Leukemia	6.8	4.1	166
Urinary cancer	7.2	5.1	141
Arteriosclerotic heart disease	256.4	190.4	135
Respiratory cancer	18.0	16.1	112
Cirrhosis of liver	10.2	9.1	112
Peptic ulcer	5.9	5.4	109
Digestive cancer	52.1	55.1	95
Motor vehicle accidents	22.9	27.3	84
Diabetes mellitus	15.2	19.1	80
Accidents, other than motor vehicle	32.8	47.6	69
Vascular lesions of the central nervous system	104.7	167.2	63
Genital cancer	21.7	34.8	62
Pneumonia	24.0	49.7	48
Hypertensive heart disease	40.2	114.1	35
Tuberculosis	7.3	25.4	29
Whooping cough	0.2	1.2	17
Syphilis	1.6	10.2	16
Homicide	2.3	23.8	10

* Rates are annual mortality rates per 100,000 population, standardized as in Table 9.

Racial differences other than those between whites and Negroes have received less attention, mainly because comparable data on other racial groups have not been available. There have, however, been special studies, for example, of the mortality from cancer among Japanese and Chinese groups living in the United

States (Smith, 1956a, b). Other studies of differences between the United States and some countries of Asia with respect to specific disease frequencies are discussed in connection with the descriptive variable "place" (see Chapter 10).

Nativity

When people migrate from their country of birth, patterns of diet, occupation, recreation and so on tend to be maintained for a variable length of time in the new country. Country of birth is therefore a valuable objective characteristic which can be used to distinguish within the same community groups whose environmental patterns and inherited characteristics may differ from those of other immigrant groups and from the nativeborn. Certain of these environmental characteristics may be maintained in more or less diluted form for one or more subsequent generations, so that nativity of parents or of grandparents may also be used to distinguish groups with different environmental experiences.

In the United States, full use has not been made of the rich material available in this connection. In part this is due to the lack of sufficiently detailed population data. For example, in recent censuses data have not been published for small demographic units on the age distribution of the foreignborn population according to specific country of birth. This considerably limits epidemiologic use of this variable, since immigration from specific countries to the United States tended to be concentrated in certain periods of time and to involve particular age groups; statistical comparisons of nativity groups that do not consider the question of age are usually quite meaningless. Much of the work in this area has therefore been done by special studies rather than by the collation of regularly reported data. A well-known example is the demonstration by Zinsser (1934) that cases of typhus occurring in New York City between 1910 and 1933 were almost exclusively limited to persons born in Eastern Europe where typhus was endemic. This observation led to the concept of typhus occurring as the recrudescence of a long-latent infection.

In New Zealand it has been shown that death rates from lung cancer were higher among persons born in England than among those nativeborn. This difference suggests the possibility that

early environment may play some role in establishing later lung cancer risk (Eastcott, 1956).

Malzberg (1940) has studied rates of admission to mental hospitals in relation to country of birth for the population of New York State. Admission rates were higher among foreignborn than among nativeborn persons, and were higher for those born in Ireland or Scandinavia than for the three other large foreignborn groups (Italian, German, English). There were striking differences in this pattern between the several major diagnostic groups of mental illness. This study was made possible by the provision of special census data on the age distribution of the foreignborn population.

In the United States, disease rates have been compared in the nativeborn and foreignborn populations without respect to specific country of birth of the foreignborn. As an indicator of environmental experiences associated with ethnic group such examinations are of limited value, since the foreignborn population is extremely heterogeneous with respect to ethnic background, and the total experience of the foreignborn may be indicative merely of the relative proportion of persons of specific nativity in the studied populations. However, particularly in studies of mental disease rates, such examinations have been of value in suggesting the possible influences common to people who have changed their country of residence. This question is considered further in the discussion of migration as a variable per se (see Chapter 11).

Religion

Ethnic background and religious belief are so strongly correlated that religion may be used as an indicator of ethnic group. For example, persons of Irish and Italian ancestry in the United States are predominantly Catholic, and those of British and Scandinavian ancestry predominantly Protestant. The more recent the ancestral migration, the stronger this association. In most American communities, however, both Catholic and Protestant groups include such a variety of ethnic backgrounds that these religions may be of little practical use for identifying groups that have any degree of ethnic homogeneity. The Jewish religion, on the other hand, does appear to identify a group that, perhaps be-

cause of a common ethnic background (in most American communities the Jewish group is predominantly Eastern European in ancestral background), and perhaps because of a closer influence of religion on everyday practices, is more homogeneous with respect to certain characteristics (e.g., circumcision, moderate use of alcohol and tobacco, and occupation) than the population at large. There is evidence that, at the present time, the Jewish religion also identifies a greater degree of genetic homogeneity than exists in the American population as a whole. Thus religion is capable of distinguishing groups with respect to rates of certain diseases and characteristics known to be genetically determined — for example, Tay-Sachs disease and ABO blood group. It would seem likely that the usefulness of religion as an epidemiologic variable in this particular context may be expected to diminish rapidly in succeeding generations.

On the other hand, some less common religious groups have established small social and reproductive units that show no sign of integration with other groups and may maintain if not increase their homogeneity. These groups, however, are usually of a local nature similar to those discussed in a later section of this chapter (page 116).

While "religion" may be the actual characteristic used to identify population groups, it is clear that explanations for differences in disease rates may or may not be sought in some feature of religious practice. On the one hand, specific methods of birth control, circumcision, and abstinence from tobacco are features respectively of Catholics, Jews, and Seventh Day Adventists that may be thought of as part of religious practice. All have been put to epidemiologic use. On the other hand, prior to the study of Zinsser mentioned previously (page 112), typhus in New York City was thought of as characteristically a disease of Jews. It had no more to do with the Jewish religion than the circumstance that the majority of the immigrants from Eastern Europe to New York City at that particular time were Jewish; nevertheless, the observation that most of the affected individuals were Jewish led to the eventual correct explanation of the occurrence of cases in an area where typhus was not known to be endemic.

A great deal of interest in religious differences has focused in the field of cancer, probably because of the well-known infre-

quency of cancer of the cervix among Jewish women. This infrequency has been noted among Jewish women in many different countries and economic groups. An examination of mortality rates from cancer in New York City according to religion indicated that cervical cancer was not the only form of cancer to show striking differences according to religion (Table 17). There are few appreciable differences between the rates for Catholics and Protestants evident in the table. As noted, these religions contain a variety of ethnic groups, and an unusual cancer experience of one ethnic group might not be revealed in such an examination. However, rates for the Jewish group show comparatively high levels for lymphosarcoma, leukemia, glioma, melanoma, and cancers of the kidney, pancreas and large intestine; and low

TABLE 17

Estimated death rates from selected malignant neoplasms in three religious groups in New York City, 1955*

Site	Males			Females		
	Catholic	Jewish	Protestant	Catholic	Jewish	Protestant
Cervix uteri	—	—	—	20.9	8.5	21.7
Tongue	8.7	2.1	8.3	—	—	—
Buccal and pharynx	15.3	4.5	12.4	—	—	—
Esophagus	21.1	12.9	21.3	3.2	5.9	5.2
Larynx	18.0	7.2	16.9	—	—	—
Lung	49.8	34.3	55.7	7.7	7.5	6.9
Stomach	61.8	76.4	75.9	40.9	48.5	39.9
Rectum	41.9	39.9	45.0	30.4	26.7	25.9
Prostate	41.8	30.0	49.8	—	—	—
Bladder	24.2	23.8	30.8	10.1	7.0	13.5
Ovary	—	—	—	22.4	30.7	27.5
Lymphosarcoma	5.4	14.6	6.8	4.4	9.0	4.4
Leukemia	18.2	19.7	13.1	8.8	18.9	18.2
Glioma	3.8	7.4	5.1	3.6	5.8	3.8
Melanoma	2.3	5.9	2.4	—	—	—
Kidney	10.6	18.2	15.1	4.9	8.4	4.1
Pancreas	27.8	38.4	27.9	18.7	26.3	19.0
Large intestine	55.9	83.8	57.2	59.1	77.0	58.9
All malignant neoplasms	585.1	594.6	638.4	437.3	510.9	467.7

* Rates are annual rates per 100,000 population aged 45 years and over.

rates are seen for neoplasms of the tongue, upper respiratory tract, esophagus and prostate, as well as for cervical cancer.

Local Reproductive and Social Units

Small local populations that have maintained environmental and some genetic homogeneity through inbreeding have of course been the subject of intensive anthropologic study, but have been little used for epidemiologic purposes. Perhaps this relates in part to the difficulty of assessing frequencies of any but the most common diseases in groups of such small size. However, there is increasing interest in the addition of accurate medical data to anthropologic studies. Preliminary investigations of certain North American Indian tribes, for example, show disease patterns so strikingly different from those of North American whites that closer study seems indicated (Hadley, 1955; Smith *et al.*, 1956).

Within the white populations of Europe and the United States a few such localized groups exist. For example, there have been studies of specific disease rates among isolated inbreeding Alpine populations. Most of these, however, have been interpreted with particular reference to genetic hypotheses. Religious belief is also occasionally associated with the formation of local groups which maintain varying degrees of inbreeding. One of the most extreme of such groups is perhaps the Hutterites, who have maintained an almost completely isolated genetic and social unit for three centuries. There are only fifteen family names in the entire sect. The unique characteristics of this group have been utilized in a recent study of the frequency and nature of mental disorders among them (Eaton and Weil, 1955).

INTERPRETATION

Associations with descriptive variables are sought primarily for the purpose of developing hypotheses concerning the role of specific experiences in disease causation. In interpreting associations with descriptive variables, therefore, the objective is frequently not to find the explanation of the association, but rather to suggest a number of different possible explanations. The former objective, while it might be desirable, is usually unattainable. As noted in Chapter 4, substantial hypotheses rarely come

from knowledge of an association with one variable. But, if associations with two or three variables are known, each one of which may have a number of explanations, hypotheses may frequently be derived by noting which explanations are common to all (or to most) of the noted associations.

The following is suggested as a simple check list to be used in seeking possible explanations of descriptive associations. It is a check list rather than a classification, since any one explanation may conceivably appear in several of the categories. It applies to most descriptive variables, although it is discussed here in the context of sex and ethnic group.

(1) Errors of measurement
(2) Differences between groups with respect to more directly associated demographic variables
(3) Differences in environment
(4) Differences in bodily constitution
(5) Differences in genetic constitution

Errors in measurement may result in apparent associations at several levels. For example, there may be differences between groups in access to medical care and hence diagnosis. Minor illnesses may be more frequently diagnosed in certain groups of males than in females because of ready access to industrial medical care facilities. United States national statistics indicate lower rates of congenital malformation for Negroes than for whites, but a considerably greater proportion of deliveries of Negro infants occurs in rural areas of the South where medical care (and hence observation) is inadequate. Even when facilities are equally available to all groups, there may be differences in utilization. For example, women are more reluctant than men to undergo physical examination. Even in the same medical care plan, utilization rates tend to be lower for Negroes than for whites, and for Catholics and Protestants than for Jews (in the United States).

Differences in precision of diagnosis must also be considered, even if access to and utilization of medical care can be considered equal. Thus, venereal disease is more difficult to diagnose in the female than in the male. Certain skin rashes may be more difficult to diagnose in Negroes than in whites.

Even if rates of diagnosis are equal, reporting may vary. For example, in morbidity surveys in which anamnestic data are sought by household interview it is most commonly a female who is interviewed, and the reporting of data for females in such households is therefore likely to be more complete than that for males. Differences in reporting frequencies also occur between ethnic groups, for example, in connection with the reporting of minor degrees of mental disturbance.

Errors of measurement commonly associated with the various sources of descriptive data have been discussed elsewhere (see Chapter 6). The understanding and consequent reduction of such errors comprises a major part of the research program of the U. S. National Health Survey (1958).

Differences between groups with respect to more directly associated variables are most obvious in relation to age differences, which are usually present between sex or ethnic groups. It was noted in Chapter 7 that age variation in disease rates is usually so marked that it is impossible to examine associations with other variables unless the data are adequate to eliminate the possible effect of age differences between the compared groups. This may be done by examination of age-specific rates for the compared groups or by the use of "standardized" rates, which are summary rates adjusted to allow for age differences. The methods of standardization are described in textbooks of biostatistics.

In addition to age, sex and ethnic groups may differ with respect to distribution by almost any other descriptive variable. Thus ethnic groups in the same community often differ with respect to socio-economic status and occupation. Similar procedures to those used to eliminate the effect of age may be applied. Thus, if a difference between whites and Negroes is observed, it would be logical to compare whites and Negroes within restricted ranges of socio-economic status and/or within restricted categories of occupation. If the differences between whites and Negroes persist within such restricted subgroups, it is unlikely that differences in socio-economic status or occupation explain the observed ethnic differences. Conversely, if a difference between occupational or socio-economic status is observed (see Chapter 9), it might be desirable to determine whether these differences persist in comparisons of persons of similar ethnic backgrounds.

Similar standardization techniques to those used for age may be applied to obtain summary rates after elimination of the effect of one or the other variable.

The desirability of identifying the descriptive variable that most closely localizes differences between groups in disease rates frequently arises in interpreting differences between ethnic groups. In particular, it is desirable to determine, if possible, whether a difference in disease incidence is associated with the particular circumstances under which the ethnic group is living at the time of observation, or whether it is a more permanent feature of the ethnic group. This may sometimes be accomplished by making the same ethnic comparison under different local conditions. For example, the fact that stomach cancer rates are much higher in Japan than in the United States may be attributed either to an innate (for example, genetically determined) susceptibility of Japanese compared to United States residents or to some environmental circumstance characteristic of life in Japan. However, Smith (1956a) has shown that while rates of stomach cancer for Japanese living in the United States are higher than rates for whites in the same country, rates for Japanese living in Hawaii are higher than for Japanese living in the United States and rates for Japanese in Japan are even higher than for those in Hawaii. Further, within the Japanese Hawaiians, rates of stomach cancer are higher in those born in Japan than in those born in Hawaii. This pattern suggests that the causal factor that appears to be characteristic of Japanese in Japan is a characterictic of a way of life that is relatively susceptible to change on migration and is certainly not of the constancy that would be associated with genetic determination.

This example demonstrates how differences first identified as characteristic of ethnic groups may later be shown to be related to rather transient features of such ethnic groups. On the other hand, the fact that many groups of Jewish women under a wide variety of circumstances have equally low rates of cervical cancer suggests that the causal factor (or its absence) is a relatively stable long-term characteristic of this ethnic group. We may note that, even if a characteristic is a stable characteristic of a particular ethnic group, it is not necessarily genetic in nature. One feature of the Jewish group that has been offered to explain plausibly the

unusual cervical cancer experience is the high rate of circumcision, a characteristic that can be considered genetic in only a most indirect way.

Differences in environment between sex or ethnic groups may be exemplified by the tremendous number of disease-related experiences associated with differences in occupation and/or socioeconomic status (see Chapter 9).

Nutritional differences between ethnic groups have received considerable attention through attempts to relate levels of arteriosclerotic heart disease in various ethnic groups to levels of intake of fat and other dietary constituents. Ethnic differences in vitamin deficiency diseases, notably pellagra, have also been shown to be due to differences between ethnic groups in dietary patterns. Dietary differences, of an as yet unidentified nature, are also suspected of being related to ethnic variation in stomach cancer. Nutritional differences also exist between the sexes and thus have a possible bearing on the higher frequency of obesity in females than in males.

Ethnic differences in personal customs, habits and recreational patterns also are productive of environmental variation. For example, ethnic differences in quantity and quality of alcohol used are well known (at least at the anecdotal level), and differences in tobacco use are probably no less striking. Such differences exist also between males and females; they are entirely capable of explaining the sex difference in lung cancer rates, and are suspected of association with sex variation in cancer of the upper respiratory and digestive tracts. Sexual practices, including contraception, show ethnic variation and may be related to ethnic differences in diseases of the reproductive organs. Likewise, a specified sexual practice or experience may affect the health of one sex differently than that of the other.

Differences in bodily constitution associated with variation in disease rates include anatomic, physiologic, and mental (attitude) differences. Anatomic features are perhaps more important in connection with sex than with ethnic differences. Apart from the obvious differences in organs with specifically sexual functions, many more subtle anatomic differences between the sexes may have pertinence to differences in disease incidence. For example, the higher frequency of arthritis of the hip and pelvic joints in

females has been attributed to skeletal differences and consequent differences in the lines of stress imposed by weight bearing. Although less common, such differences do occur between ethnic groups. For example, the protection against the sun's rays conferred by skin coloring presumably is associated with differences noted between ethnic groups in skin cancer frequency.

Physiologic differences between males and females are illustrated by hormonal differences, suspected of association with several important diseases including atherosclerosis and several forms of neoplasm. More obviously, specific diseases of the endocrine glands (e.g., the thyroid gland) are likely to be influenced by the total hormonal environment in which the gland is operating. Again, physiologic differences between ethnic groups are less obvious than those between the sexes. The level of immunity at the time of challenge with organisms is one type of physiologic characteristic which varies between ethnic groups. For example, the low frequency of paralytic poliomyelitis in Negroes in the United States is attributed to frequent early exposure (associated with overcrowding and other features of low socio-economic status) at an age when paralysis is rare, resulting in subsequent high levels of immunity.

Mental or attitude differences as explanatory of differences in disease rates between sex and ethnic groups are most obvious with respect to the several forms of disability due to violence. (Figure 15). Here again, quite different explanations may be sought for the existence of the attitude differences in the first place. Levels of disease perception, as illustrated by sex and ethnic differences in pain tolerance, may also be thought of as part of the contribution of attitude differences to disease rates.

Differences in genetic constitution may contribute to differences in disease rates insofar as genetic constitution determines many of the more directly associated attributes mentioned above, such as, for example, skin color or the presence or absence of specific sex organs.

In addition, single major genes (that is, single genes which under suitable circumstances are associated with a demonstrable characteristic of the persons carrying them) may be transmitted in association with either sex or ethnic group. In the case of transmission of such genes in a particular sex pattern, the pattern is

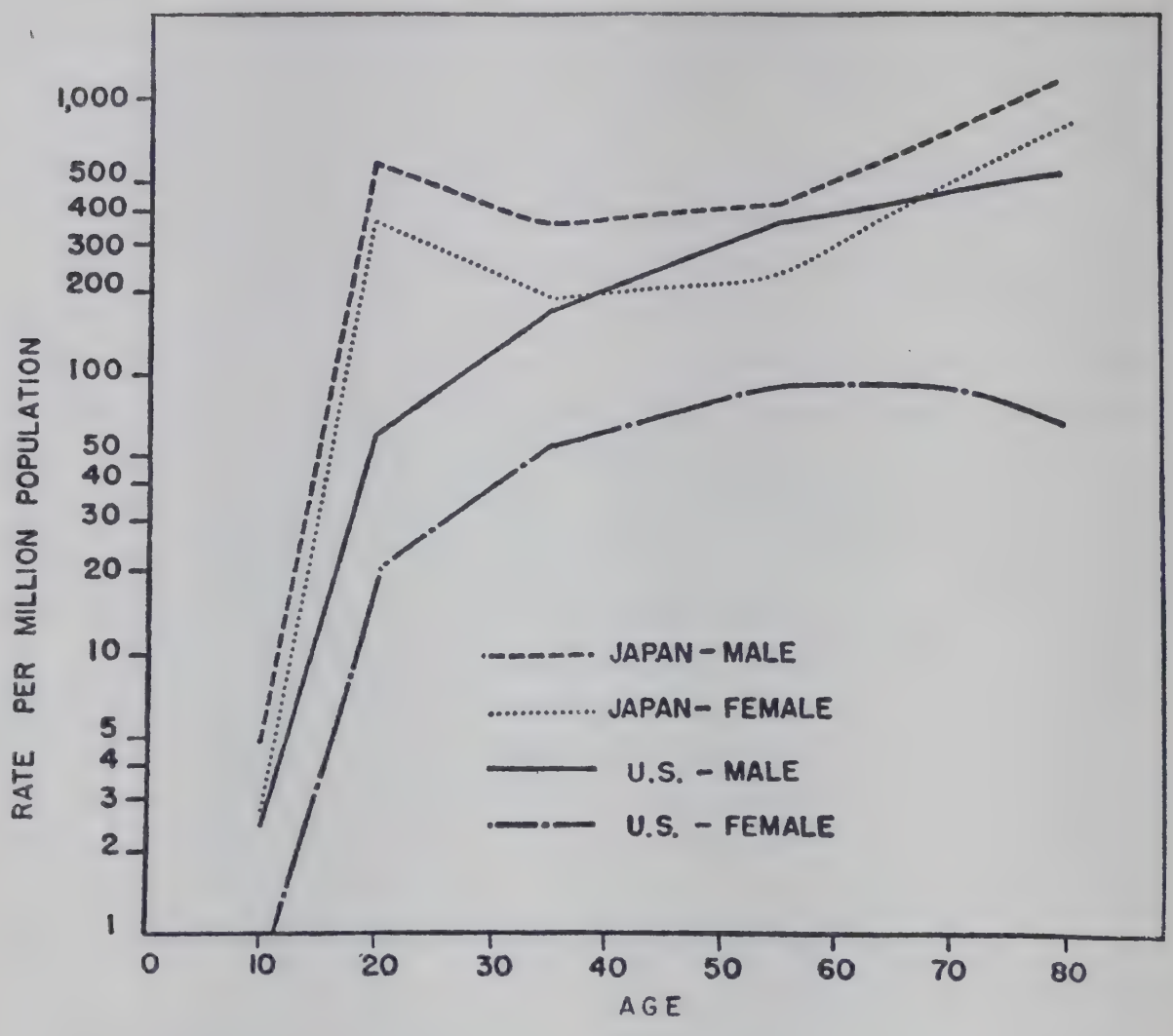


FIGURE 15. Age-specific death rates attributed to suicide in Japan and in the white population of the United States, 1954 to 1956. (Data from the World Health Organization, 1958.)

attributable to disease-associated genes being located on the same chromosome pair that determines sex ("sex-linkage"). Examples include hemophilia, certain forms of muscular dystrophy and red-green color blindness. Sex-linked diseases are usually readily identifiable when major genes of a high degree of manifestation are involved.

Ethnic concentration of genetically determined disorders, on the other hand, is usually due to limitation of mating to the particular ethnic group over many generations, and the consequent limitation to that ethnic group of genetic material which originally arose as new mutations in that group or was introduced to it by immigrants. When major genes are involved, as in the case of sickle-cell anemia, thalassemia and Tay-Sachs disease, genetic analysis will usually identify this type of ethnic concentra-

tion. However, greater difficulty arises either when manifestation of a major gene is weak and dependent on other unknown factors, or when many minor genes are involved, none of which alone has a demonstrable effect. Evidence for this type of determination of ethnic differences is difficult to obtain. Simultaneous consideration should be given to collateral evidence of genetic determination *within* ethnic groups (familial incidence, twin studies, etc.), showing that the characteristic rate (whether high or low) is a stable feature of the ethnic group (see page 119), and demonstration that other types of explanations (discussed above) are not sufficient to explain noted differences.

9

Other Characteristics of Persons

At this juncture, it should be restated that the variables most commonly used in descriptive epidemiology are mainly the ones for which large quantities of data are routinely assembled by various agencies for their own particular purposes. In other words they are selected because of their ready accessibility and do not necessarily represent ideal types of epidemiologic information. It also is well to note again that those variables being discussed are spoken of as "characteristics of persons" only for convenience, and not because associations with them are deemed to reflect etiologic factors intrinsic within the individual.

OCCUPATION

Purposes and Interpretation

Occupation, which fortunately forms a part of many routinely collected records, is used as a descriptive epidemiologic variable for a number of rather different purposes. These include the following:

(1) Measurement of social or economic status. Occupation is one of the basic indices of the elusive concept "socio-economic status."

(2) Identification of specific risks associated with exposure to noxious agents peculiar to certain occupations. This is the classic epidemiologic use of occupation that has resulted in the identification of the detrimental effects of soot, lead, benzene and aniline derivatives, ionizing radiation, and numerous other chemical and physical agents.

(3) Measurement of the effects of different conditions of work. The general conditions under which an occupational group works, quite apart from exposure to specific physical and chemical agents, may play a part in determining its disease experience. While associations dependent on such circumstances may be less strong than those dependent on specific chemical or physical risks, they may be of even greater epidemiologic interest, since their relevance to circumstances outside the particular occupational group may be far reaching.

These three types of purpose also suggest general areas within which speculation as to the interpretation of a noted occupational difference in disease rates might lie. However, before assuming an interpretation as specifically related to occupational conditions as suggested by any of these three areas, one must consider the alternate types of explanation referred to in Chapter 8, particularly those included in "errors of measurement," and "differences between groups with respect to more directly associated variables." Errors of measurement may clearly influence occupational comparisons when perception, diagnosis, or reporting of illnesses differ between compared groups. Differences between occupational groups with respect to more directly associated variables arise because of the marked influence of variables such as sex, age and ethnic group on choice (or availability) of occupation. In addition to demographic variables such as these, characteristics even more directly indicative of health status, such as temperament, physique and physical handicaps, play a major part in the selection of occupation, and must be considered in the interpretation of statistical associations between occupation and health status.

Occupation as an Index of Socio-economic Status

By far the largest body of data relating to occupation and disease derive from the surveys of the Registrar General for England

and Wales, in which deaths in census and adjacent years have been related to the occupational distribution of the population as determined by the census. These surveys have been undertaken periodically since 1851, the latest reports being for the years 1930 to 1932 (1938) and 1950 (1954). The reports have covered both occupied and retired males aged 16 and over, and beginning with the reports for 1930 to 1932 have also dealt with single occupied women and with married women classified according to the occupations of their husband.

Occupations have been divided into five broad classes — Professional (I), Intermediate (II), Skilled (III), Partly Skilled (IV) and Unskilled (V). In 1950 certain subclasses of classes III, IV, and V were also examined. In the analysis of this material, in addition to deriving age-specific mortality rates in each social class, considerable use has been made of the standardized mortality ratio (the number of deaths, either total or cause-specific, in a given occupational group expressed as a percentage of the number of deaths that would have been expected in that occupational group if the age-and-sex-specific rates in the general population had obtained). These ratios for death from all causes and for certain causes that manifest marked socio-economic gradients are shown in Table 18. Mortality generally increases with decreasing social class, although quite divergent trends are seen for different diseases. The irregularity of the trends in the data for

TABLE 18

*Standardized mortality ratios for males aged 20 to 64 in five occupational classes, England and Wales**

Occupational ("social") class	All causes			1950				
	1921-1923	1930-1932	1950	Tuberculosis	Bronchitis	Gastric ulcer	Coronary heart disease	Leukemia
I (Professional)	82	90	97	64	33	56	153	150
II (Intermediate)	94	94	86	62	53	81	101	110
III (Skilled)	95	97	102	103	97	97	107	104
IV (Partly skilled)	101	102	94	95	103	99	81	79
V (Unskilled)	125	111	118	149	172	144	88	89

* From the Registrar General's Decennial Supplements (1938, 1954).

1950 may be because they so far represent material based only on a sample survey of the population and on deaths for 1 year. A larger study covering mortality during 5 years is planned.

Occupation of father may also be used to classify infants according to socio-economic status. Infant mortality rates derived in this way are shown in Table 19. Infant mortality increases sharply with decreasing socio-economic status, and although total rates have fallen markedly between 1921 and 1950, the relative disparity between socio-economic groups has not changed.

TABLE 19

*Infant mortality rates per thousand related to occupational class of father**

Occupational class of father	Aged less than 1 year			1950	
	1921	1930-1932	1950	Aged under 4 weeks	Aged 4 weeks to 1 year
I (Professional)	38.4	32.7	17.9	12.9	4.9
II (Intermediate)	55.5	45.0	22.2	16.2	6.0
III (Skilled)	76.8	57.6	28.1	17.6	10.5
IV (Partly skilled)	89.4	66.8	33.7	19.8	13.9
V (Unskilled)	97.0	77.1	40.7	21.9	18.8
Total	79.1	61.6	29.3	18.1	11.2

* From the Registrar General's Decennial Supplements (1938, 1954).

Special Risks

Standardized mortality ratios for specific occupations have also been presented in the Registrar General's Decennial Supplements, and have given valuable leads in the identification of specific risks that might be associated with the physical and chemical exposures of certain occupations. There have also, of course, been innumerable special studies of occupational risks, to the extent that the area now constitutes a discrete medical specialty. The methods are in general similar to those of cohort studies of other special groups exposed to supposedly noxious circumstances, and are dealt with later (see Chapter 13). Hueper (1952) has presented a comprehensive review of neoplastic disorders known to be associated with occupation, and also (1950) detailed methods for occupational cancer surveys.

Conditions of Work

The variations in conditions of work that occur both between and within occupations have received increasing attention in the light of a number of recent studies, particularly of coronary heart disease and mental disorders. For example, Morris and his colleagues (1953) have shown that the comparative frequencies of coronary heart disease in several occupations suggest lower frequencies and less severe disease in active than in sedentary occupations. Persons in different levels of authority in the same industry have been compared with respect to hypertension and arteriosclerosis (Lee and Schneider, 1958) and to neurosis (Fraser *et al.,* 1947). Friedman *et al.* (1958) have studied the blood cholesterol levels of accountants during periods of occupational stress, noting significant increases during such periods. Cairns and Stewart (1951) have related tuberculosis rates to conditions of work within a section of the British shoe industry.

SOCIO-ECONOMIC STATUS

Measurement

Association of disease frequency with socio-economic status, as with other descriptive variables, is measured by stratification of the population according to arbitrary levels of the variable and derivation of disease frequency rates in the several subgroups so formed, with due allowance for age, sex, and racial variations between the populations of the subgroups. A difficulty in this case is that socio-economic status is a theoretical concept still awaiting clear specification of meaning. So many variables, such as occupation, family income, living conditions, social prestige, and so on, are encompassed that, in practice, one variable which can be objectively defined is commonly used as an indirect indicator for epidemiologic purposes. Since each variable may measure a different component of the socio-economic complex, it is not surprising that associations with "socio-economic status" may differ according to which measure is used. It would seem more objective to describe such associations as associations with the particular variable used

(e.g., income), rather than as associations with the general concept of which the selected variable is only a part.

Occupation. The use of occupation in this context has already been described. As noted, this variable can be used to stratify retired persons (on the basis of past occupation) and married women and children (on the basis of the occupation of the family wage earner), as well as the wage earners themselves.

Family income. Income might be thought of as a more direct measure of the concept of socio-economic status than is occupation. The use of this variable is limited, however, by the lack of availability and comparability of numerator and denominator data for the computation of rates. Although data on income are collected for the population in the decennial censuses, they are not so generally available for ill persons as are data on occupation. Furthermore, even when income data are available for sick persons, as, for example, in hospital records, it is questionable whether the method of ascertainment is sufficiently similar to that of census data to allow meaningful comparisons. Family income has been most satisfactorily used in morbidity surveys based on household interviews, when data on income and other general characteristics can be assembled for the population at the same time that its health status is being ascertained. Data from the National Health Survey are shown in Figure 16. For both sexes (although more so for males than females) the average number of days of bed-disability decreases regularly and strikingly with increasing income.

Various measures of income have been used. These include total family income, average family income per person in the household, and more complicated measures such as average family income per "adult unit," children being counted as certain percentages of an "adult unit." Actually, all these measures are highly correlated and the choice of one instead of another is usually not critical; the measurement of disease frequency is not so precise as to be affected by the minor variations caused by application of the different methods of computation of income.

Area of residence. For most cities, census data are available that allow classification of census tracts or other small administrative units of the city according to some measure of socio-economic status. Thus, census tracts may be classified according to the me-

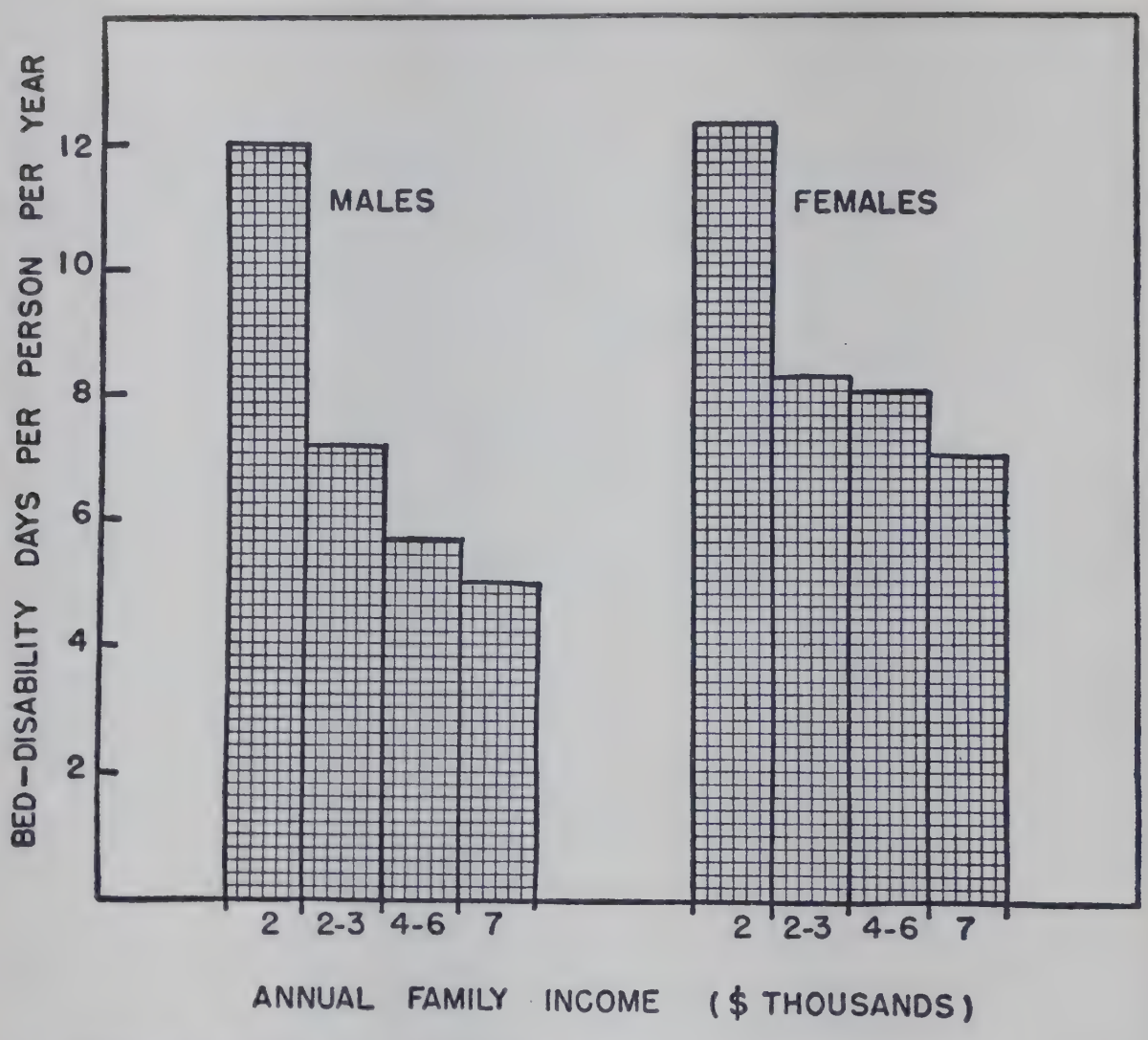


FIGURE 16. Number of bed-disability days per person per year related to sex and family income in the United States, July 1957 to June 1958. (Data from the National Health Survey, 1959.)

dian income of their populations, the per cent of the population with incomes below a certain level, the per cent of the population in certain occupational categories, or measures of housing standards such as rental or degree of crowding. Criteria of greater intimacy have included the proportion of servants to those served, and the proportion of houses with baths. Disease frequency rates may be computed and compared for residents of areas classified according to one of these indices. Although the socio-economic status of the ill persons is not known individually, it is inferred from their area of residence. Data based on first admissions to mental hospitals from the City of Birmingham, England, between 1950 and 1952 are shown in Figure 17. The several wards of the city were classified according to the percentage of dwellings lacking certain amenities (Lowe and Garratt, 1959).

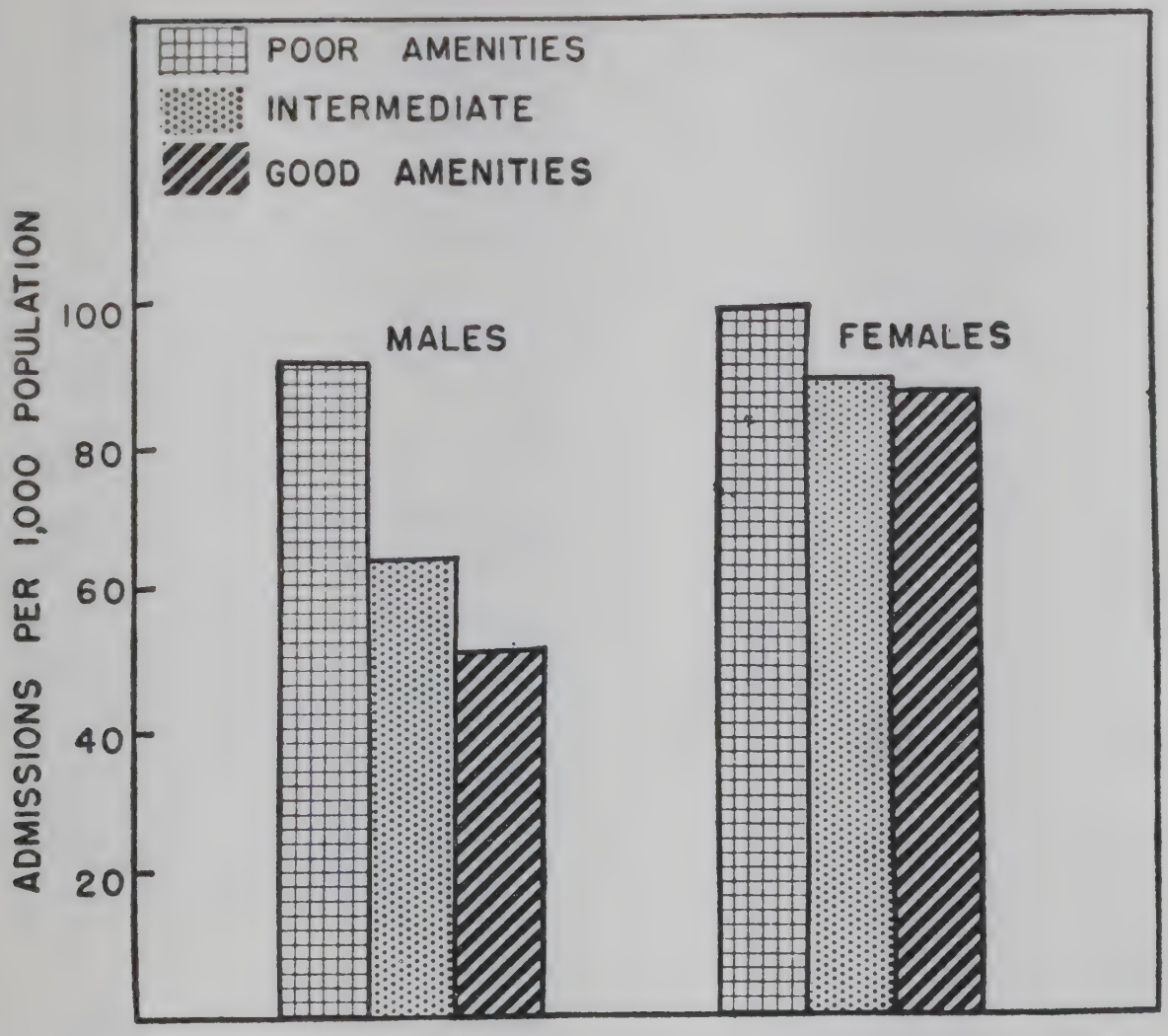


FIGURE 17. Average annual first admission rates to mental hospitals related to area of residence, Birmingham, England, 1950 to 1952. (Data from Lowe and Garratt, 1959.)

The reliability of this method varies greatly from one city to another and may vary also within the same city from time to time. While there is no question as to the tendency for the residences of populations of low and high economic status to cluster in certain parts of cities, the administrative areas for which the necessary census data are available may not be small enough to discriminate clearly between such clusters. If differential disease rates are demonstrated by this method, then (subject to the usual precautions) a socio-economic association can probably be assumed. On the other hand, if no association is demonstrated, two alternative explanations must be considered: (1) that the disease is not associated with socio-economic status, or (2) that the classification of areas is not sufficiently discriminating to detect a socio-economic disease association, which in fact may exist. For example, Martin (1955) showed that the correlation between certain

indices of overcrowding and infant mortality in the twenty-nine Boroughs of London was very high at the beginning of this century, but was virtually nonexistent in the years 1950 to 1952. Since other indices, such as occupation (see Table 19), suggest that the correlation between infant mortality and socio-economic status is still high, it is likely that the change noted by Martin is due to a change in the characteristics of the London boroughs, rather than in the association of infant mortality with socio-economic status.

Combination of variables. Occasionally, an investigator will devise an index combining a number of variables by an arbitrary formula to fit in with his concept of socio-economic status. The index might be based on a weighted combination of such items as income, rental, family size, occupation, education and residence. The latter three, for example, were used by Hollingshead and Redlich (1958) in their study of mental disorders. While such an index has usefulness in an individual study, it has the disadvantage of rarely being used in its same form by other investigators, and therefore hindering comparisons of results of several studies.

Other indices. Various other indices have been used and are appropriate for particular situations. For example, there have been comparisons of mortality among taxpayers and nontaxpayers (Chapin, 1924), and among the families of British peers and the general population (Bailey and Day, 1861). During the last century, comparisons of disease frequencies in army officers and private soldiers were popular.

Interpretation

The many different environments experienced by persons of different socio-economic levels must be considered before explanations of noted variations in disease frequency are proposed. These include access to and use of facilities for medical care, education and other cultural characteristics, nutrition, crowding, and special circumstances related to the spread of microbiologic agents. An example that helped to elucidate the etiology of pellagra is shown in Table 20.

Before one speculates on possible causal mechanisms it is necessary, as in the case of other descriptive variables, to consider er-

rors of measurement, the influence within socio-economic groups of other variables such as ethnic group, and the question whether occupancy of a low socio-economic group may itself have been determined by incipient or actual disease occurrence rather than the other way round. This possibility is, of course, important in social systems in which it is difficult to undergo long-term illness without bankruptcy; socio-economic associations with rates of prevalence or death from chronic illnesses should invariably be examined with this possibility in mind. Associations based on incidence rates will be less affected, particularly if the illness involved is one without prolonged prodromata.

TABLE 20

Rates of pellagra related to family income in seven cotton-mill villages of South Carolina, 1916 *

Half-month family income per "adult male unit" ($)	Both sexes			Crude rate per 1000	
	Number of persons	Rate per 1000			
		Crude	Age-standardized	Males	Females
< 6	1312	42.7	41.0	30.8	54.4
6-7	1037	26.0	24.8	11.5	40.7
8-9	784	12.8	14.2	10.7	14.7
10-13	736	4.1	5.2	0.0	8.0
14 +	291	3.4	2.5	6.2	0.0
Total	4160	23.3	23.3	14.9	31.6

* From Goldberger *et al.* (1920).

Even when costs directly associated with the illness (e.g., the costs of medical care) have not determined the socio-economic levels measured in the study, there may be difficulties of this type if the disease under study has prodromal or preclinical stages that might affect a person's earning ability. For example, considerable controversy has centered around the question of whether the higher frequencies of diagnosed mental disorders that have been observed in lower socio-economic groups truly results from a higher frequency of occurrence among these groups or from a social "drift" that might be expected of persons with incipient undiagnosed mental disorders. There is no ready solution to this question, although with a decreasing interval between the obser-

vation of the difference and the onset of manifest incapacity the uncertainty should correspondingly become less.

MARITAL STATUS

General Mortality

Unlike socio-economic status, marital status is a variable which is easy to measure; ready stratification of the population may be made into four major groups — single, married, widowed and divorced. The problem of ensuring similar definition of these terms in both the affected individuals and the population to which they are referred for the derivation of rates is an important though not difficult one.

While marital status appears routinely on death certificates and in census records, data on deaths in relation to marital status are not published routinely in this country. They have, however, been the subject of special reports of the National Office of Vital Statistics for the years surrounding the censuses of 1940 and 1950 (1945, 1956). Age-standardized rates for 1949 to 1951 are shown in Table 21. They demonstrate a relationship that is characteris-

TABLE 21

Mortality rates related to marital status in the white population aged 20 or more, United States, 1949-1951†*

Marital status	Total	Male	Female
Single	15.6	20.6	10.8
Married	11.3	13.1	8.7
Widowed	14.2	21.4	12.1
Divorced	22.5	26.7	13.7
Total	12.6	15.1	12.4

* Rates are average annual mortality rates per 1000 population. Standardization is to the age distribution of the total United States population aged 20 and over in 1950.

† Data from National Office of Vital Statistics (1956).

tic — the lowest mortality being for married persons and the highest for widowed and divorced. This relationship is seen for both sexes.

Table 22, also calculated from age-specific rates given in the

1950 report, shows that the relationship between the rates for single and married persons is similar for most of the major causes of death.

TABLE 22

Mortality rates from selected causes in single and married white persons aged 20 and over, United States, 1949-1951†*

Cause of death	Males		Females	
	Single	Married	Single	Married
Respiratory tuberculosis	96.7	22.7	25.0	11.4
Tuberculosis, other forms	3.4	1.3	1.8	0.8
Syphilis	14.8	6.6	2.0	2.2
Malignant neoplasms:				
Total	260.0	207.1	470.4	182.5
Buccal and pharynx	13.7	7.0	1.6	1.8
Digestive organs	115.7	87.6	73.0	65.5
Respiratory system	44.5	33.5	7.8	7.2
Breast	0.7	0.4	54.0	33.2
Genital organs	23.7	25.0	40.0	38.2
Urinary system	15.5	13.3	6.6	5.8
Leukemia	8.0	8.8	5.6	5.8
Diabetes mellitus	21.9	18.3	16.5	29.0
Vascular lesions of C.N.S.	182.1	138.9	136.6	124.7
Rheumatic heart disease	28.7	18.6	23.8	17.4
Arteriosclerotic heart disease	525.6	409.9	212.0	179.1
Chronic nephritis	33.0	23.1	20.2	18.4
Influenza and pneumonia	73.5	27.0	34.1	18.3
Peptic ulcer	21.5	11.8	3.8	2.4
Motor vehicle accidents	61.9	33.7	11.0	10.8
Other accidents	99.2	46.1	36.4	21.1
Suicide	43.7	20.5	7.7	6.2

* Rates are average annual rates per 100,000 population, standardized as in Table 21.

† Data from National Office of Vital Statistics (1956).

Three major explanations of such a relationship are:

(1) Persons in poor health, or in the presymptomatic stages of ill health, tend to remain single.

(2) Persons who "live dangerously," and are consequently exposed to a wide variety of disease-producing agents and situations, tend to remain single.

(3) There are differences in the ways of life of single and married persons that result from the presence or absence of the married state and that are causally related to certain diseases.

It is likely that all three of these explanations play a role in some of the noted relationships. Studies of cancer and mental disorders serve to illustrate some of the possibilities.

Cancer

Table 22 shows that a number of sites of cancer exhibit frequency differences between married and single persons, but interest so far has largely focused on cancers of those organs of the female most directly involved in reproductive and other marital functions. It is desirable, if possible, to separate the effects of reproduction or fertility from other concomitants of the married state. If the data are available, this can be done by comparing single women with infertile married women. An example is shown in Figure 18. It is seen that for carcinoma of the cervix rates for married infertile women are almost twice as high as those for single women. Carcinoma of the corpus uteri, on the other hand, appears about equally frequent in these two groups, although married women as a whole would show lower rates than single women because of the lower rates evident for fertile women. Similarly, it has been shown that married women have lower rates of breast cancer than do single women; this in part can be attributed to a lower rate of breast cancer among women who have borne children, but it is also evident when single and married infertile women are compared.

It is likely that the associations of carcinoma of the cervix and of the breast with marital status are both explicable in terms of marital function. That cervical cancer is extremely rare in nuns and that rates show an inverse relationship with age at marriage, independent of fertility, suggest that frequency of sexual intercourse may be the variable responsible for the descriptive association with marital status. Lilienfeld (1956) has shown that artificial menopause confers some degree of protection against breast cancer, and that married women more frequently have the menopause induced artificially than do single women. This relationship is also independent of differences in fertility, and may explain the difference between single and married infertile women in the frequency of breast cancer (MacMahon and Feinleib, 1960).

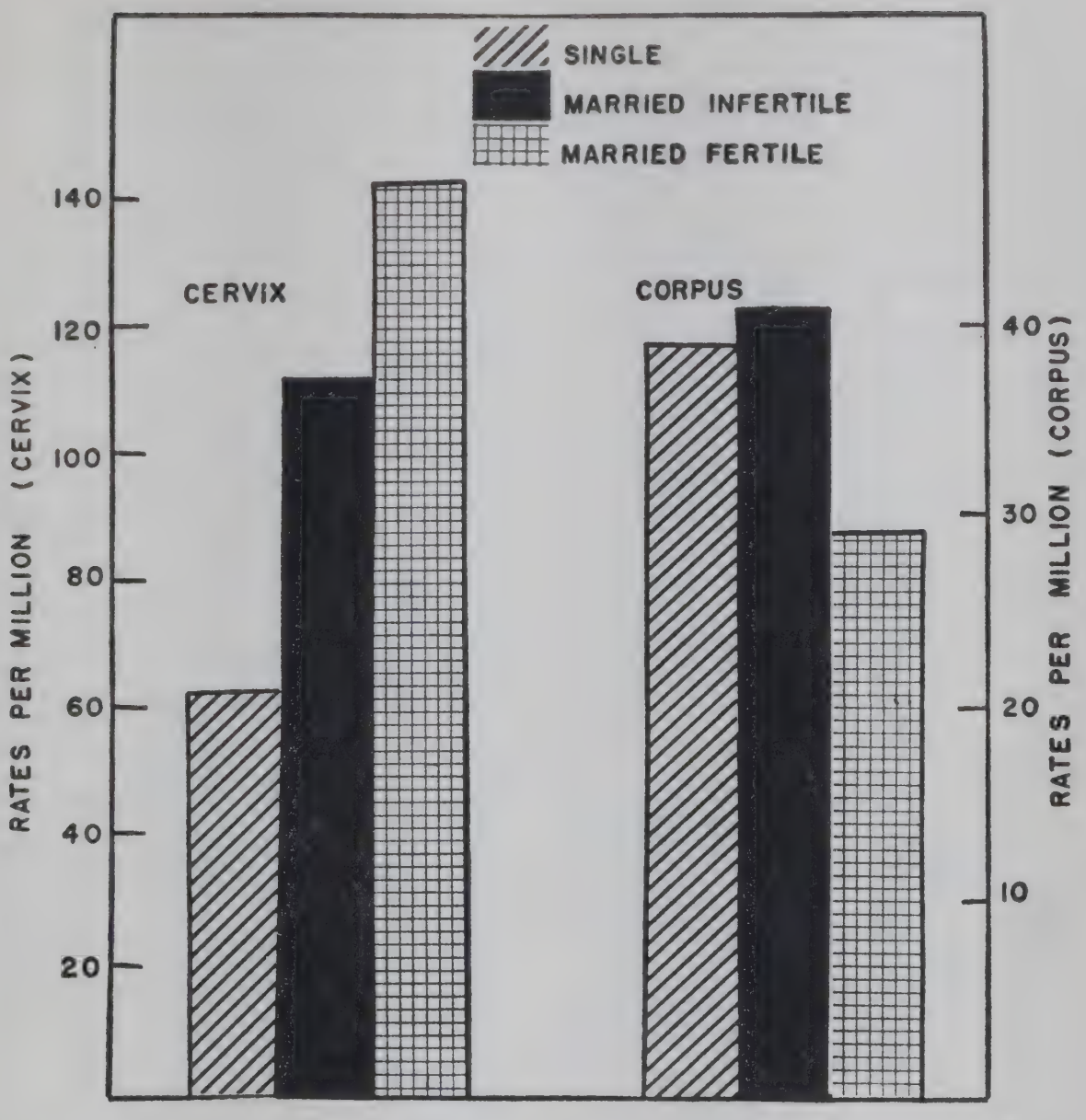


FIGURE 18. Equivalent average death rates from cancer of the uterus related to marital status and fertility in England and Wales, 1948 to 1949. An equivalent average death rate is the sum of the age-specific death rates divided by the number of age groups summed. It is the equivalent of standardization to a standard population which has equal numbers in each age group. In the present case, data are restricted to females aged 15 or more. (From Logan, 1953.)

Mental Disease

Point prevalence rates of persons in mental hospitals in the United States in 1950 are shown in Figure 19. The most striking difference is between the rates for single and married persons. In addition to the magnitude of the difference, there is an interesting reversal of the sex pattern between the two marital status groups. Among single persons, rates for males are higher than for

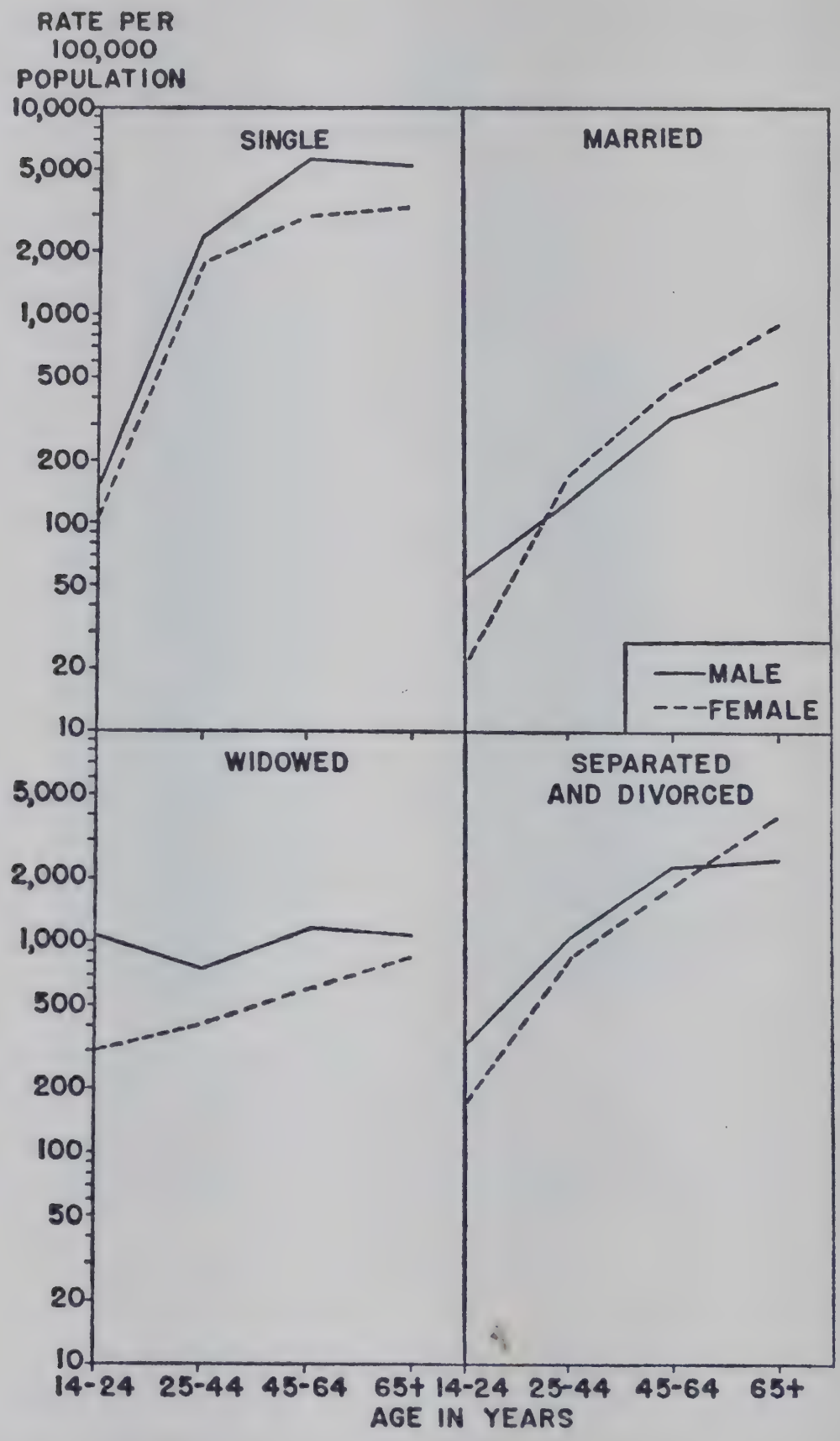


FIGURE 19. Point prevalence of patients in mental hospitals per 100,000 population according to age, sex and marital status, United States, 1950. (Data from U. S. Bureau of the Census, 1953b.)

females, yet among married persons the converse is true for all age groups except the youngest (14-24 years). Similar relationships can be demonstrated in incidence (first admission) data.

These relationships have a variety of possible explanations, as previously suggested. If the explanation is based on the premise that persons with manifest or incipient mental disorders tend to remain single, then the effect of this selection must either be greater for males than for females or there are additional etiologic influences which affect single males and married females. An observation that is difficult to explain on the basis of selection of marital status is the generally higher rates of mental disorders for widowed than for married persons, particularly in the younger age groups. The most likely explanation of this relationship seems to be that the stress of bereavement or of widowhood has a causal relationship to the manifestation of mental disorder.

Still a further explanation of the difference between the rates for married and single persons has to do with the use of mental hospital data as estimates of the frequency of mental disorder. It is possible that the actual frequencies of disorder are the same but that single persons are hospitalized more frequently because of the absence of alternative care. Here again, the high rates for widowed persons do not seem to conform to such a hypothesis, since it is difficult to believe that alternative accommodation would be less readily available to young than to old widows and widowers.

Concordance between Marital Partners

Another area of investigation deals with the possible influence of a common environment on the health of marital partners, studies being conducted to determine whether husband and wife tend to have the same disease experiences. Ciocco (1940) found among the spouses of persons who died from tuberculosis, pneumonia, heart disease or cancer a higher than expected number who died of the same condition. In morbidity data, Downes (1947) found that pairs of husbands and wives both had a chronic disease significantly more frequently than would be expected by chance; there was a significant association of the same illness in both spouses for hypertensive vascular disease and arthritis.

Such studies are useful in pointing to disorders in which causal

environmental circumstances may be sought in experiences that are common to husbands and wives. The possibility that marital partners are consciously or unconsciously selected with respect to characteristics that will determine the development of similar diseases (e.g., ethnic group, or physical attributes) must of course be given consideration.

FERTILITY

Rates for a few disorders among females have been related to the number of children borne. Disorders of pregnancy provide the most obvious of such relationships. Toxemia, placenta previa and uterine efficiency, for example, show marked associations with maternal parity. Mechanical disorders of the reproductive tract, such as uterine prolapse, also show relationships to maternal fertility.

Examples of a less immediate relationship are that rates of cervical cancer increase and that those of breast cancer decrease with increasing parity. Wynder (1955) and others have suggested that the association of cervical cancer with parity may be secondary to associations of each with age at first marriage, but no satisfactory explanation of the inverse relationship between breast cancer rates and parity has been substantiated.

Although there has been some work on time of occurrence of psychosis in relation to child birth (see Chapter 11), no data are available on the risk of mental disorders in relation to number of children borne. With the great amount of information on the fertility of the female population which is collected by the censuses of most countries, it is surprising that epidemiologic use of this descriptive variable has been virtually restricted to studies of uterine and breast cancer. In addition, if there were census data available for male fertility, it would seem reasonable to examine male disease patterns with respect to fertility.

10

Place

The importance of variation in the frequency of different disease manifestations from place to place has long been recognized. It has formed the basis of much reasoning leading to hypotheses of etiology. In addition, knowledge of the geographic distribution of disease has obvious utility for administrative purposes.

METHODS

Early knowledge of the geographic distribution of disease was entirely qualitative, and derived from the sporadically published observations of medical practitioners. Hirsch's three-volume *Handbook of Geographical and Historical Pathology,* translated into English in 1883, is a gargantuan compendium of such observations and is regarded by many as the beginning of modern epidemiology. Because of its reliance on qualitative data, perhaps it should instead be regarded as the full flowering of anecdotal epidemiology rather than as the beginning of a new approach.

The quantitative knowledge of present-day epidemiology is based in part on special surveys and in part on analysis of routinely collected and published vital data. The use of routinely collected statistics of causes of death and notifications of infectious diseases has the advantage of allowing exploratory examinations

of a large number of diseases to determine the likelihood of significant geographic variation. When probable differences are demonstrated, special confirmatory surveys may be instituted. However, special surveys are expensive and unlikely to be undertaken unless routine data suggest specific hypotheses which need testing.

On the other hand, routine data have a number of serious disadvantages. First, there are the difficulties consequent on the variations that exist from place to place in standards of medical care, diagnosis and reporting of illness or death. Second, there is the virtual restriction of such data to infectious diseases and disorders reported as causes of death. Third, the limits or regions for which the data are reported are usually administrative, and the geographic factors influential as determinants of disease frequency may not be so accommodating as to approximate administrative boundaries; consequently, failure (with routinely available data) to demonstrate geographic differences in the distribution of a disease does not rule out the possibility that place may indeed play a role in its occurrence.

According to the scale of the map, we may distinguish several types of comparison that are commonly undertaken. It should be clear, of course, that these comparisons must be considered in the light of the effects of differences in age and sex composition between the compared populations.

International Comparisons

Data from many nations relating to causes of death and notifications of infectious diseases are currently collected and periodically published by the World Health Organization. Although the published data are restricted to those countries from which reporting is believed to be most adequate, the difficulties associated with differences in diagnosis and reporting are more serious in comparisons of data between nations than those associated with most other geographic data. Uncertainty whether even gross differences may be due to such artifacts is felt particularly when they are found in areas outside of Europe, North America and Australia. Nevertheless, the gross differences that have been demonstrated, and even those that are less striking, should not be written off as inconsequential simply because of this uncertainty.

They will no doubt be susceptible to increasingly critical examination as reporting procedures improve.

Most striking, of course, are differences between tropical and temperate areas in the frequency of major parasitic diseases. In such situations, where both clinical opinion and statistical material indicate that certain diseases that are epidemic in particular areas are rare in others, there can be little doubt that the differences are real. Less obviously real, and therefore all the more provocative, are the unusual distributions observed for certain noninfectious diseases. Japan, for example, with relatively low rates of death from coronary artery disease, has very high rates of hypertension and cerebrovascular accidents; unusually high death rates from stomach cancer exist in Germany and Iceland, as well as in Japan; Great Britain is worst off with respect to lung cancer, although the United States, while not in quite the same class, is not very far behind. If the United States needs honors it can get them in the fields of coronary artery disease and leukemia, although for the latter the distinction must be shared with Denmark and Sweden. Perhaps the main value of statistical data such as these is to provide a mechanism for the broad classification of countries into those with probably high, probably average and probably low rates of an individual disease; the data appear too fragile to undergo a finer classification at the present time. Nonetheless, they do constitute the most important source of information on international variation in disease frequency available at the present time and deserve serious attention.

Special studies of international variation are well illustrated by those in which the purpose has been to correlate coronary artery disease frequencies with the dietary constituents in various parts of the world. Special clinical studies have been made among groups of people in areas selected because of observed dietary patterns (see, for example, Scrimshaw *et al.*, 1957). Valuable information on the distribution of atherosclerosis and on the levels of serum cholesterol has been obtained, although there is still no consensus as to which dietary constituents are most closely correlated with the variation in coronary disease frequency (Yerushalmy and Hilleboe, 1957). Particularly interesting in such studies are attempts to rule out ethnic differences as being explanatory of observed geographic differences. For example,

Miller and his colleagues (1958) found that the serum cholesterol levels of men whose parents were born near Naples but who themselves had lived all their adult lives in the United States, while comparable to those of other Americans, were significantly higher than those of native Neapolitans examined by Keys *et al.* (1954).

In many international studies of this type, information both on the collateral variables (such as diet) and on health and disease characteristics (such as cholesterol level) is obtained from the same persons. In this type of study the fact that the group observed may not be representative of the population residing in the area is not crucial. However, if disease rates derived from such special study groups are to be compared with characteristics which are representative of the population of the area, the selection of a representative sample in the special study group does become critical and considerably increases the technical difficulties which must be faced.

Variation Within Countries

Just as death statistics and notifications of infectious diseases constitute the major source of knowledge of variation in disease frequency between countries, so they are the major sources of data for comparisons within countries. Again, use must be made of the administrative areas for which deaths and population bases are reported. In general, the smaller (geographically) the administrative areas examined, the more accurately the boundaries of areas of high or low disease frequency can be drawn, provided that the individual areas have populations large enough to yield reliable estimates of disease frequency.

For example, Figure 20 shows standard mortality ratios for deaths from leukemia in each state of the United States, while Figure 21 shows data calculated for individual "economic subregions," the latter being smaller population groups that incorporate several contiguous counties within states. At a distance, the impression obtained from the two figures is similar — both, for example, indicate low frequencies in the Southeast and a pocket of high frequency in the North Central part of the country. However, when examined more closely the maps show some significant dissimilarities. For example, Figure 20 suggests that

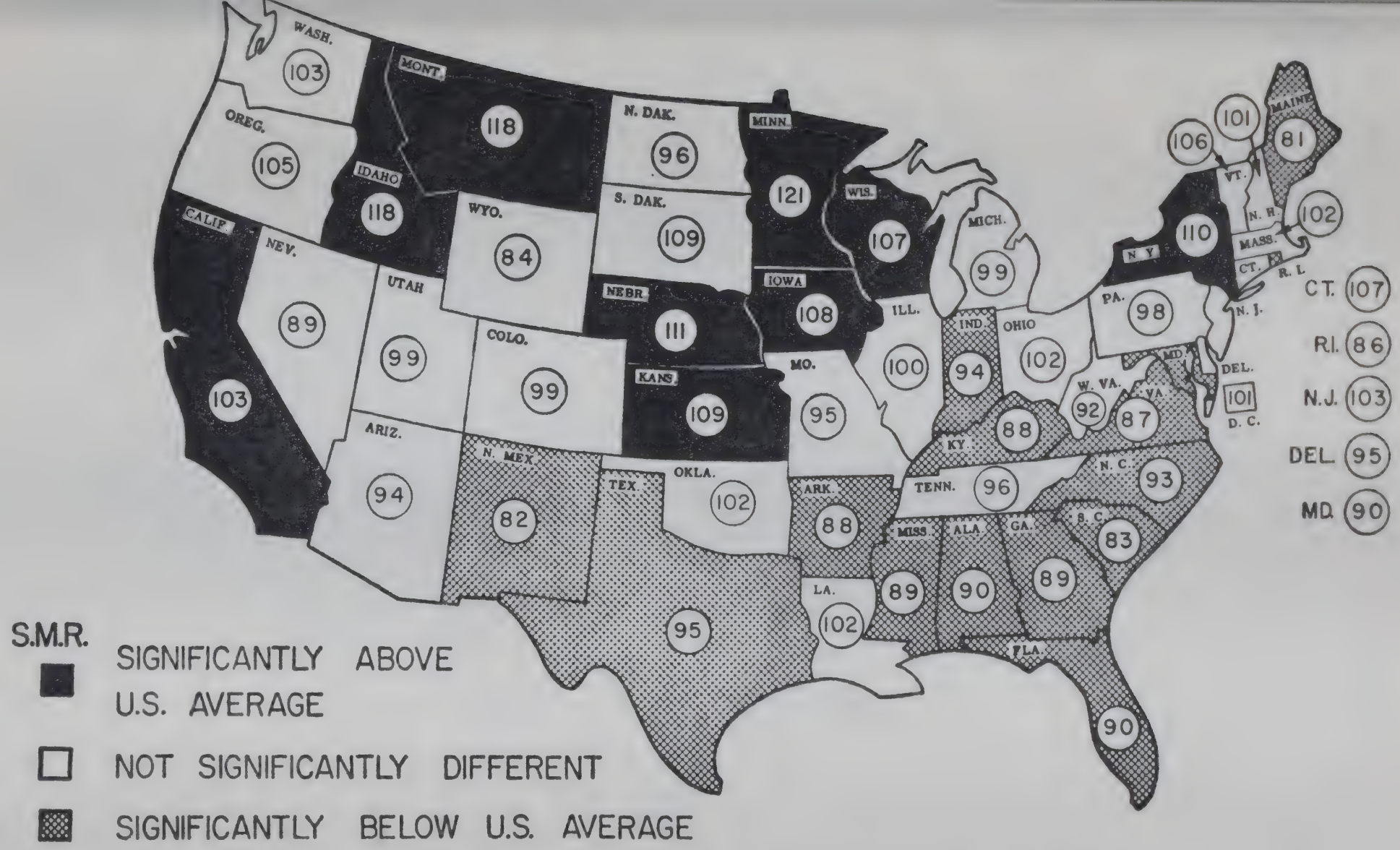


FIGURE 20. Standard mortality ratios (S.M.R.) for deaths from leukemia in each State, U.S.A., 1949-1955. The standard mortality ratios take into account the distribution of the population of each state according to sex, race, and age. (Reproduced, with permission, from Gilliam, 1960.)

rates are very high for Minnesota and that an area of high rates extends across Wisconsin to the shore of Lake Michigan. But in Figure 21 it is seen that the highest frequencies actually are found in Western Minnesota, extending into parts of the Dakotas. Similarly, Figure 21 shows that the high rate for New York State compared with surrounding states (seen in Figure 20) is in fact a characteristic of the New York Metropolitan Area, but not of upstate New York.

A detailed geographic delineation of disease frequencies may not always be required. There may, for example, be interest primarily in the question of whether there is a difference in frequency between the northern and the southern parts of a country. Even so, an examination of data for small administrative units will help to clarify whether a difference noted between the north and the south (when examined as two large units) reflects a difference of the areas in general or is due to the presence within one or the other of geographic localizations of very high or low frequency. The presence of a few pockets of high frequency in one-half of a country would obviously require a different interpretation than the finding of a widespread higher frequency in one-half compared to the other, the latter finding, for example, being more meaningful in terms of the dichotomy north and south.

Classification of places within a country need not be made on a spatially contiguous basis. For example, it may be desirable to compare counties or states having large agricultural populations with those having small farming populations. The same possibility — that one or the other category might show a high disease frequency rate because it includes one or two areas of very high frequency (because of an unsuspected reason) — can be examined by comparison of areas classified according to levels of the relevant variable rather than according to a simple geographic dichotomy, or by use of a two-way table arranging varieties of areas into those having rates above and below average levels of disease frequency.

The examination of a geographic variation may go beyond a description of the levels of variation observed to the development and exploration of hypotheses attempting to explain the noted variation in terms of other characteristics known about the regions.

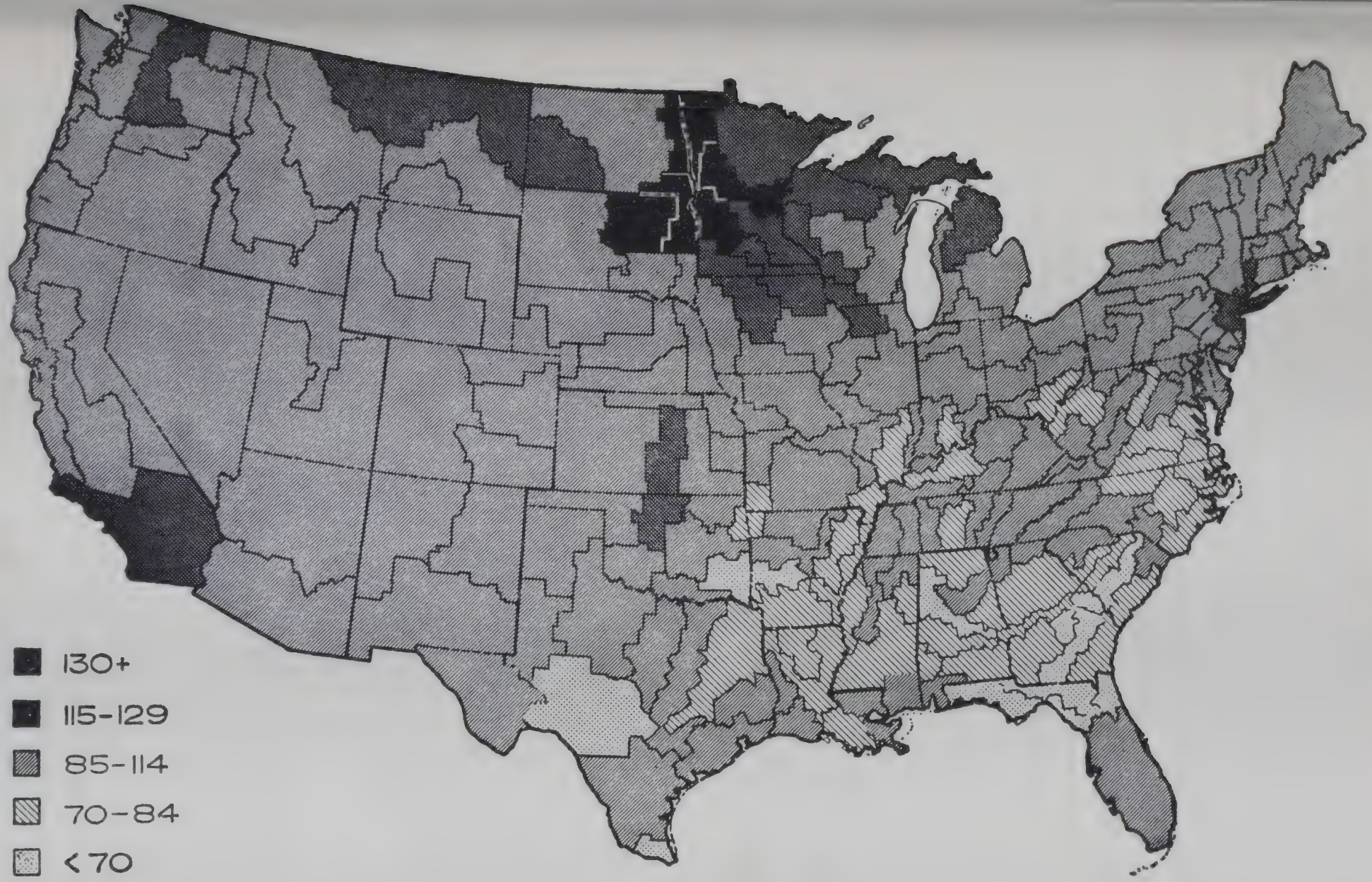


FIGURE 21. Standard mortality ratios for deaths from leukemia for economic subregions of the United States, 1949-1951. The standard mortality ratios take into account the distribution of the population of each state according to sex, race, and age. (Reproduced, with permission, from Gilliam, 1960.)

This further exploration may involve comparison of the disease frequencies with the frequencies of those characteristics suspected of being relevant, by means of correlation scatter diagrams or by more formal methods of mathematical correlation. An example (Figure 22) is to be found in the work of Dean (1946) and others in the identification of a major determinant of the geographic variation in frequency of dental caries — the fluoride content of drinking water.

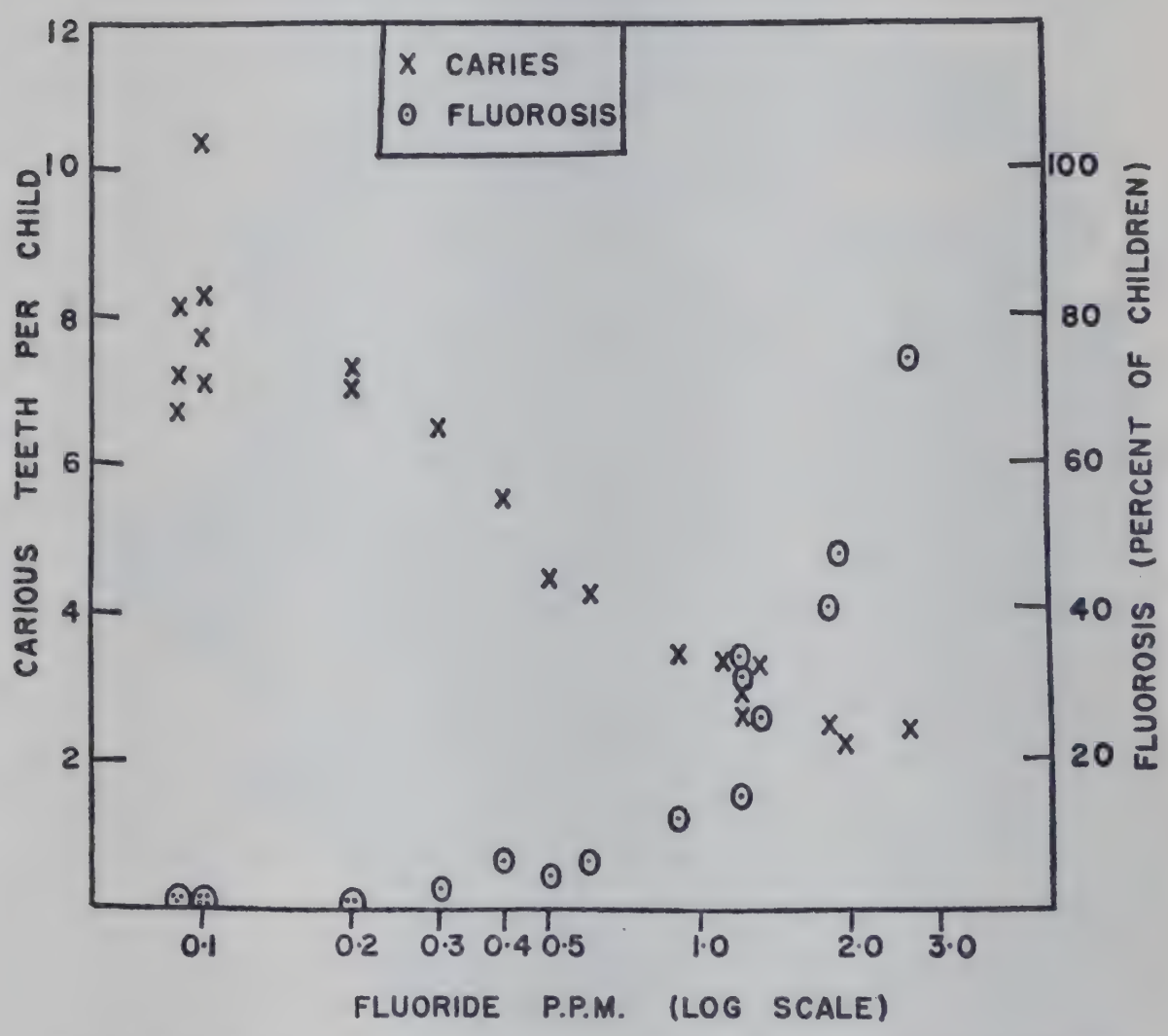


FIGURE 22. Fluoride content (parts per million) of drinking water correlated with dental caries experience and with evidence of any degree of dental fluorosis (including "very mild") in white school children aged 12 to 14 years in twenty-one cities of four states. (Data from Dean, 1946.)

Such information of course is not derived from routinely processed data but from special surveys designed to examine particular hypotheses. Special surveys both of animals and of man are frequently undertaken to develop an understanding of the distribution of infectious diseases. In man such surveys may in-

volve search for existent clinical cases of the disease. Frequently, however, they are concerned mainly with evidences of residua of infections — for example, the presence of antibodies in the serum, trachoma scars or splenic enlargement. Clearly, examination of populations for scars and other leftovers from past infections is more rewarding from the standpoint of numbers found affected than is a search for current cases. It should be noted, however, that the information derived refers to a peculiar form of point prevalence — the frequency of having had and survived a past infection. The evaluation of such a rate is necessarily difficult.

A social phenomenon that involved the clinical examination of certain age groups of men from all over the United States and was consequently of great convenience to the geographically minded epidemiologist, if to nobody else, was the mobilization of the Armed Forces during World War II. For example, several studies of the prevalence of dental caries in relation to area of residence of inductees were undertaken, a summary of which is shown in Figure 23. Similarly, Gentry *et al.* (1955) used data from the same source to examine the area of residence of inductees with sarcoidosis. The latter study provides a good example of the relationship between ethnic group and geography. Since sarcoidosis occurs much more frequently among Negroes, the finding that most cases came from the southeastern part of the United States at first sight was not surprising. However, even when allowance was made for the geographic distribution of the Negro population, there still appears to have been a geographic concentration of the disease in this area and to certain smaller areas within it.

Urban-Rural Comparisons

A special type of intracountry comparison that does not involve classification into contiguous areas is the examination of differences in disease frequency according to degree of urbanization. Increasing use of this variable has demonstrated that economic and industrial development is not an entirely unmitigated blessing.

Both the U. S. National Office of Vital Statistics and the Registrar General for England and Wales publish tabulations of cause of death according to urban or rural residence. In the United States the division is along county lines, counties being classified

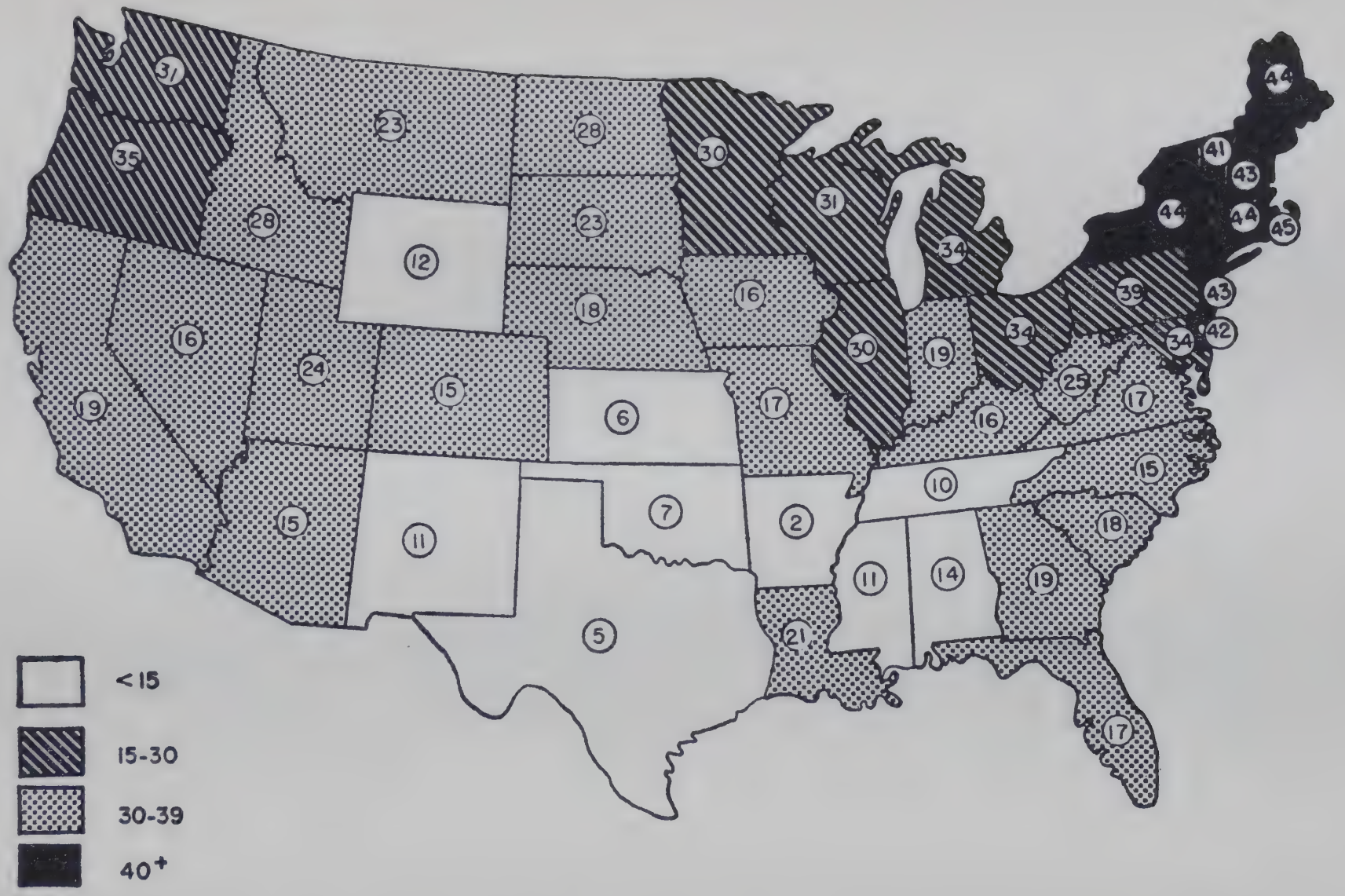


FIGURE 23. Mean rank of states with respect to dental health in three large studies of military inductees. Measures of "dental health" were different in the three studies, but each of the measures used was indicative primarily of the past or present prevalence of dental caries. (From Dunning, 1953.)

according to whether they lie within the Standard Metropolitan Areas established by the Bureau of the Census ("metropolitan counties"), or not ("non-metropolitan counties"). A special tabulation of deaths for the years 1949 to 1951, undertaken in connection with the Air Pollution Medical Program of the U. S. Public Health Service, provides age-standardized death rates for 102 causes of death in 163 Standard Metropolitan Areas (Manos, 1957). In addition to allowing comparison of the metropolitan areas with the rest of the country, these data allow comparisons of the individual metropolitan areas. A distinction is also drawn between counties with a central city in a Standard Metropolitan Area and counties that are peripheral to a large city although still included in a Standard Metropolitan Area. The latter are presumably less urbanized than the former.

Data for England and Wales published by the Registrar General provide even more detail for study of urban-rural differences. Tabulations of deaths by cause, age and sex are given for six large "conurbations" (coalescenses of major towns), urban areas outside the conurbations divided into three groups according to population size, and rural areas.

Table 23 illustrates urban-rural comparisons for selected causes of death as derived from the data for 163 Metropolitan Areas in the United States. Death rates are generally higher for urban than for rural areas, both for all causes of death and for many individual causes. The data for deaths from "symptoms, senility and ill-defined conditions" indicate that diagnostic and reporting difficulties may enter into these comparisons, although, at least in the United States and Great Britain, the effect of such possible differences is small. The most striking excess rates in urban, by comparison with rural, areas is seen for arteriosclerotic heart disease, tuberculosis and cirrhosis of the liver. The excess rural death rates from nephritis and vascular lesions of the central nervous system in non-metropolitan areas, although small, are of interest because they are contrary to the general trend.

Urban-rural comparisons are also frequently made during the course of special studies. For example, in the infectious diseases, inferences regarding mechanisms of spread have been made from comparisons of urban and rural areas with respect to frequency, age distribution and rapidity of spread of the disease. Diseases

MP 100 N60
0541(
COMMUNITY HEALTH C
LIBRARY
AND
DOCUMENTATION
UNIT

TABLE 23

Mortality rates for selected causes of death, by degree of urbanization, United States, white population, 1949-1951†*

Cause of death	Males			Females		
	Metropolitan counties		Non-metropolitan counties	Metropolitan counties		Non-metropolitan counties
	With central city	Without central city		With central city	Without central city	
All causes	1161	1061	1049	777	743	722
Arteriosclerotic heart disease	338	306	246	165	152	113
Cancer (all sites)	166	156	123	139	136	118
Vascular lesions of C.N.S.	98	96	107	92	93	98
Accidents	74	71	99	32	31	38
Hypertension	60	56	53	60	60	52
Tuberculosis	30	19	21	11	8	12
Other infectious diseases	14	13	19	8	8	14
Nephritis	16	16	20	12	13	17
Diabetes mellitus	14	13	12	21	21	17
Cirrhosis of liver	17	13	8	8	7	4
Peptic ulcer	10	9	7	2	2	1
Symptoms, senility and ill-defined conditions	9	8	18	5	5	12

* Rates are average annual rates per 100,000 population, standardized to the age distribution of the total population of the United States in 1950.

† Data from Manos (1957).

spread by contact and respiratory infections are acquired later in life and often less frequently in the country than in the city. On the other hand, intestinal infections and many diseases transmitted from animals to man have greater opportunity to affect human beings in rural areas.

Results of two cancer morbidity studies, one a special survey in Iowa and the other of data from the Danish Cancer Registry, are shown in Table 24. Minor differences in this table must be interpreted with caution, since, as noted, diagnostic standards may well be somewhat higher in urban than in rural areas, even within these two limited geographic regions. However, strikingly higher rates for urban areas are seen in both studies for cancers of the pharynx, esophagus, lung, bladder and cervix.

TABLE 24

Urban-rural differences in incidence of cancer of selected sites, Denmark, 1948-1952, and Iowa, 1950†*

Site and sex	Denmark			Iowa	
	Copenhagen	Provincial towns	Rural areas	Urban area	Rural areas
Lip (M)	2.9	4.7	8.1	17.3	15.7
Tongue (M)	1.4	1.0	0.3	1.6	1.1
Pharynx (M)	1.9	0.8	0.3	4.3	1.5
Esophagus (M)	8.0	5.1	3.8	3.9	2.1
Stomach (M)	37.0	43.4	41.8	24.7	24.7
Stomach (F)	23.6	26.7	29.8	13.1	14.1
Rectum (M)	25.2	22.5	17.4	16.3	12.4
Rectum (F)	13.9	12.2	11.2	10.0	11.0
Lung (M)	43.6	15.3	8.3	29.0	10.2
Kidney (M)	8.8	4.4	3.7	4.8	3.5
Kidney (F)	5.2	3.0	3.4	3.3	2.8
Bladder (M)	15.3	7.7	5.6	20.2	12.9
Bladder (F)	4.1	2.4	2.1	6.6	3.2
Leukemia‡ (M)	8.4	7.0	6.6	14.4	11.4
Leukemia‡ (F)	6.0	5.0	4.5	10.9	7.2
Cervix (F)	38.4	34.9	20.2	43.4	23.6
Corpus uteri (F)	14.4	10.9	7.9	11.0	10.9
Breast (F)	57.4	49.3	40.9	78.0	62.4
All sites (M)	284.8	224.7	185.7	350.6	250.2
All sites (F)	272.5	241.5	208.6	351.6	261.5

* Rates are annual incidence rates per 100,000 standardized to the age distribution of the total United States population in 1950.

† From Clemmesen and Nielsen (1957) and Haenszel *et al.* (1956).

‡ For Iowa, these rates include other cancers of the hemopoietic system.

Local Distributions

Spot maps showing the distribution of cases of a disease within the local community comprise a long-established epidemiologic procedure. A classic example, Snow's (1855) investigation of the outbreak of cholera in the Golden Square area of London in August and September of 1848, is illustrated in Figure 24. The identification by Snow of the place where each of the victims of this outbreak was attacked (i.e., became frankly ill) enabled him to center attention on the common pump in Broad Street (A). He noted that the outbreak centered geographically around this pump and that the number of cases was small in areas that were

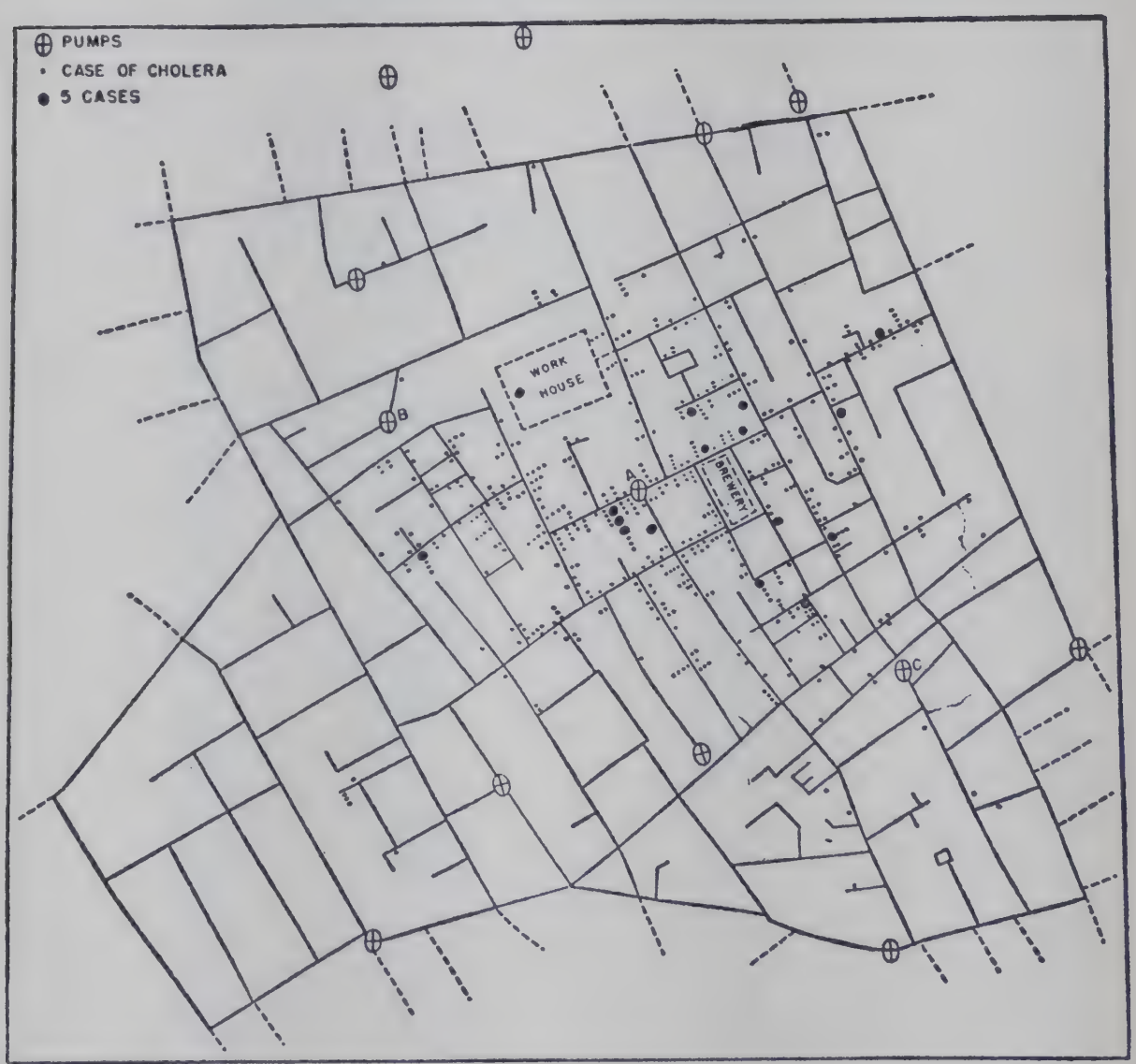


FIGURE 24. Street map of the subdistricts of Golden Square, St. James and Berwick St., London, showing the distribution of cholera cases in the outbreak of August and September, 1848. (From Snow, 1855).

closer to some other source of water supply. The localization of cases around the Broad Street pump becomes still more striking when it is noted that (1) although there are several cases in the vicinity of the pump in Marlborough Street (B) the water from this pump, being heavily and obviously polluted, was avoided by residents of this area who preferred to walk the additional distance to the pump in Broad Street; and (2) the pump in Rupert Street (C) while apparently close to several cases as the crow flies, was not close as the water-carrier travels because of the conformation of the streets in the area.

While no formal reference to any population at risk is usually included in such a map, knowledge of the distribution of the population is essential to interpretation. An area of the map may be empty of cases merely because it is not populated. For example, the 5 cases shown in the workhouse on Snow's map mean little unless it is known that the institution had 535 inhabitants. By contrast, a percussion-cap factory in Broad Street had 18 cases among approximately 200 workers. The workers in the factory used water from the Broad Street pump, but the workhouse had its own well on the grounds. Similarly, the absence of cases among workers in the brewery in Broad Street becomes of significance when it is known that 70 men were employed there. Dr. Snow, inquiring of Mr. Huggins, the proprietor, ascertained that: "The men are allowed a certain quantity of malt liquor, and Mr. Huggins believes they do not drink water at all; and he is quite certain that the workmen never obtained water from the pump in the street. There is a deep well in the brewery."

The making of a spot map such as this is commonly the first procedure in the investigation of a localized outbreak of disease. When the number of persons at risk in subsections of the affected area is large enough, the computation of rates is preferable to the spot map.

At this juncture, it is necessary to point out that comparisons of disease rates for areas as small as the ones just described can only be expected to reveal those causal factors that have operated at a time close to the onset of the disease. Man is currently so mobile, particularly in the Western culture, that he occupies during his lifetime a variety of living places. If, therefore, compari-

sons of areas are expected to reveal causal factors operative at a time distant from the onset of a disease, the use of the largest geographic units (nations) may be expected to be most revealing.

INTERPRETATION

The demonstration of an association of a disease with place implies one of the following: (1) The inhabitants of the particular place have characteristics that in themselves are of etiologic importance in the disease and that differ from those of the inhabitants of other places; (2) etiologic factors are present in the biologic, chemical and physical, or cultural environments of people inhabiting the affected places; or (3) both types of explanation are applicable.

Characteristics of Inhabitants

Characteristics of persons that produce associations with place include some of the clearly identifiable characteristics which have been discussed in the previous chapters. Age, for example, is so obvious that allowance for age differences between the populations of different places forms part of the method of the comparison and should not even enter into interpretative considerations. Mention has already been made of ethnic differences as determinants of geographic patterns. Religious groups also tend to be geographically clustered — for example, it is likely that the low death rates from lung cancer in the State of Utah can be attributed to the abstinence from tobacco required of the large Mormon population.

On the other hand, the personal characteristics involved may be more subtle and not identifiable through use of gross descriptive variables. An example of genetic variation from place to place is the variation in frequency of sickle-cell anemia in Africa. Within East and West Africa, the sickle-cell trait* tends to be most frequent in areas in which malaria rates are highest. Allison (1954) has shown that persons who are heterozygous for the sickling gene, although as frequently infected, are more resistant to the malaria parasite. It seems likely, therefore, that

* The sickle-cell trait occurs in persons heterozygous for the gene which in homozygous form is associated with sickle-cell anemia.

genetic selection in favor of persons with the sickling trait is at least partly responsible for the occurrence of high sickling rates in districts with high malaria rates (Livingstone, 1958). On the other hand, it is likely also that the present distribution of the sickling gene accounts for some geographic variation in the distribution of malaria as a disease manifestation, since sickling-trait prevalences between 20 and 30 per cent are fairly frequent among West African tribes.

Characteristics Peculiar to Place

That a noted geographic association may be explained in terms of characteristics inherent in the place of occurrence of a disease is suggested by any of the following criteria:

(1) High frequency rates are observed in all ethnic groups inhabiting the area.

(2) High frequency rates are not observed in persons of similar ethnic groups inhabiting other areas.

(3) Healthy persons entering the area become ill with a similar frequency to the indigenous inhabitants.

(4) Inhabitants who have left the area do not show high rates, and if affected at the time of emigration show signs of improvement or recovery.

(5) Other species than man inhabiting the same area show similar manifestations.

The more of these criteria that a disease satisfies, the more probably the frequency of the disease is in fact determined by a characteristic (or characteristics) directly related to place. Although all of them are rarely satisfied by a particular disease, they may be; the classic example, endemic goiter, was clearly recognized as a "place" disease long before the particular cause of its distribution was determined. The fact that one or several of the criteria are not satisfied does not rule out the possibility that strong etiologic agencies peculiar to the locality are operative in the genesis of a disease. For example, criterion (4) may be inappropriate if the disease is one with a lengthy preclinical phase or is associated with progressive and irreversible pathologic change; or criteria (1) and (5) may not pertain in the presence of such an additional association as a genetic mechanism.

It is worth noting parenthetically that differences in observed

geographic distributions may lead to erroneous as well as to enlightening conclusions when additional associations are present. For example, an important argument used against the hypothesis of an etiologic similarity between syphilis and general paralysis of the insane was that general paralysis was quite rare in many countries where syphilis was very common (Kraepelin, 1913). The reason for this discrepancy, if it is a fact, is still unknown.

If the evidence that a particular place or type of place exercises an etiologic influence on a disease seems reasonably convincing, it is possible to seek for responsible mechanisms among several of the broad (and not mutually exclusive) features which make up our concept of "place."

Biologic environment. Best understood of the features of places that contain disease-determining mechanisms are the climatic and other ecologic characteristics that determine the animal and vegetable environment within which man exists. Such characteristics influence disease patterns by providing temperature, humidity and other conditions necessary for the survival of parasites outside the body (for example, the cercariae of schistosomiasis, the larval stages of some intestinal parasitic worms, and the various leptospira); by providing conditions suitable for the reproduction of animal vectors (for example, mosquitos, tsetse flies, ticks and various ectoparasites of man and other animals); and by determining the size and kind of animal populations available for a role either as host for intermediate stages of the parasites or as reservoir of infection, for example, mollusks in schistosomiasis and the mammalian reservoirs of plague, rabies, trypanosomiasis and yellow fever).

The vegetable environment, in addition to its role in the determination of the animal environment, may have a more direct part, as illustrated by the geographic distribution of such prosaic diseases as poison ivy dermatitis and such exotic ones as ergotism.

Chemical and physical environments. Chemical conditions peculiar to certain places are capable of influencing the health of man through at least two known mechanisms — peculiarities in the chemical composition of drinking water and of inspired air. Perhaps the most geographic of all diseases — endemic goiter — has been shown to be associated with iodine lack in the soil. Sway-back in sheep is also the result of a deficiency, in this instance

of copper. The inhalation of toxic chemicals peculiar to certain localities was well illustrated by the effects of the London fog shown in Figure 1 (see page 6). It seems likely, though not fully established, that atmospheric pollution also has more gradually apparent effects on human health, including a probable role in the etiology of lung cancer. Such an association may well explain the unenviable position of Great Britain with respect to death rates from lung cancer.

Physical environment plays an obvious role in diseases and disabilities directly connected with exposure to heat, cold or lack of oxygen (altitude). Recently, there has been interest in the identification of possible variations in rates of diseases known to be associated with ionizing radiation, for example, leukemia and bone cancer, according to variation in the levels of background radioactivity that occur from place to place. Gentry *et al.* (1959) have reported evidence of such a correlation in a study of rates of congenital malformation in various areas of New York State.

Cultural environment. Social arrangements peculiar to specified places are major determinants of the biologic, chemical and physical environments to which the inhabitants are exposed. These arrangements appear to be the main forces in control of the nature and quantity of animal and vegetable species sharing the environment, of the frequency and diversity of interpersonal contact, of the presence of reservoirs and vectors for spread of infection, and of the degree and kind of physical and chemical additions to the air and water.

Social conditions determining preferences, availability and distribution of food are known to be responsible for geographic patterns in the distribution of such disorders as beriberi, pellagra and obesity. In addition, the less tangible factor of social "climate" peculiar to a time and place may be productive of manifestations that members of other social systems may view as disease, although unfortunately the evidence so far remains largely anecdotal. Hirsch (1883), for example, noted that: "Methodist evangelizing . . . spread like a pestilence from Ohio through Tennessee, Virginia, Kentucky and the Carolinas, affecting hundreds of thousands; the victims were known by various names, according to the form their enthusiasm took . . ." And Panum, during his well-known investigation of an outbreak of measles on

the Faroe Islands in 1846, having taken the opportunity to observe many other facets of life in the region, formed the opinion that both the physical characteristics of the islands and the religious attitudes of the people led to a type of mental disorder which was distinctive from disorders he had seen in other areas.

In descriptive epidemiology there is nowhere a more obvious need for the development of quantitative measures than in the area of social arrangements and relationships.

CHANGES IN PLACE

Man's increasing mobility was mentioned earlier in this chapter in connection with an epidemiologic difficulty which it engendered. However, this mobility on occasion has offered situations useful in the understanding of disease frequency as it relates to place of occurrence.

Change in place of residence and life has been used as an epidemiologic variable in the exploration of two general types of hypothesis: (1) the existing disease level of a population is affected by the entrance into it of ill (or potentially ill) persons; and (2) the health of those entering is affected by the act of migration.

The first hypothesis has stimulated a variety of interests, ranging from the role of the mass pilgrimages to Mecca in the spread of cholera to studies of the influence of introducing infected animals into laboratory populations of mice.

The second general hypothesis that the act of migration influences the health of migrants is given support not only by evidence derived from the behavior of infectious diseases, but also from that relevant to diseases which do not seem to be associated with microbiologic agents. In the instance of infectious diseases, Wheelis' work (1938), demonstrating a relationship between recent entrance into the United States Navy and high rates of communicable diseases, is particularly pertinent. For the "noninfectious" diseases well-documented support of the hypothesis is found in the area of mental disorders. Psychotic disorders appear to be more frequent among migrant groups.

This finding can be demonstrated in more than one way. The comparison of rates of nativeborn with those of foreignborn per-

sons has been mentioned in Chapter 7. Another way, involving comparison between rates for whites and non-whites, requires the additional knowledge that migration within the United States has increased at a more rapid rate for the non-white than for the white population (Figure 25). Still another involves recency of

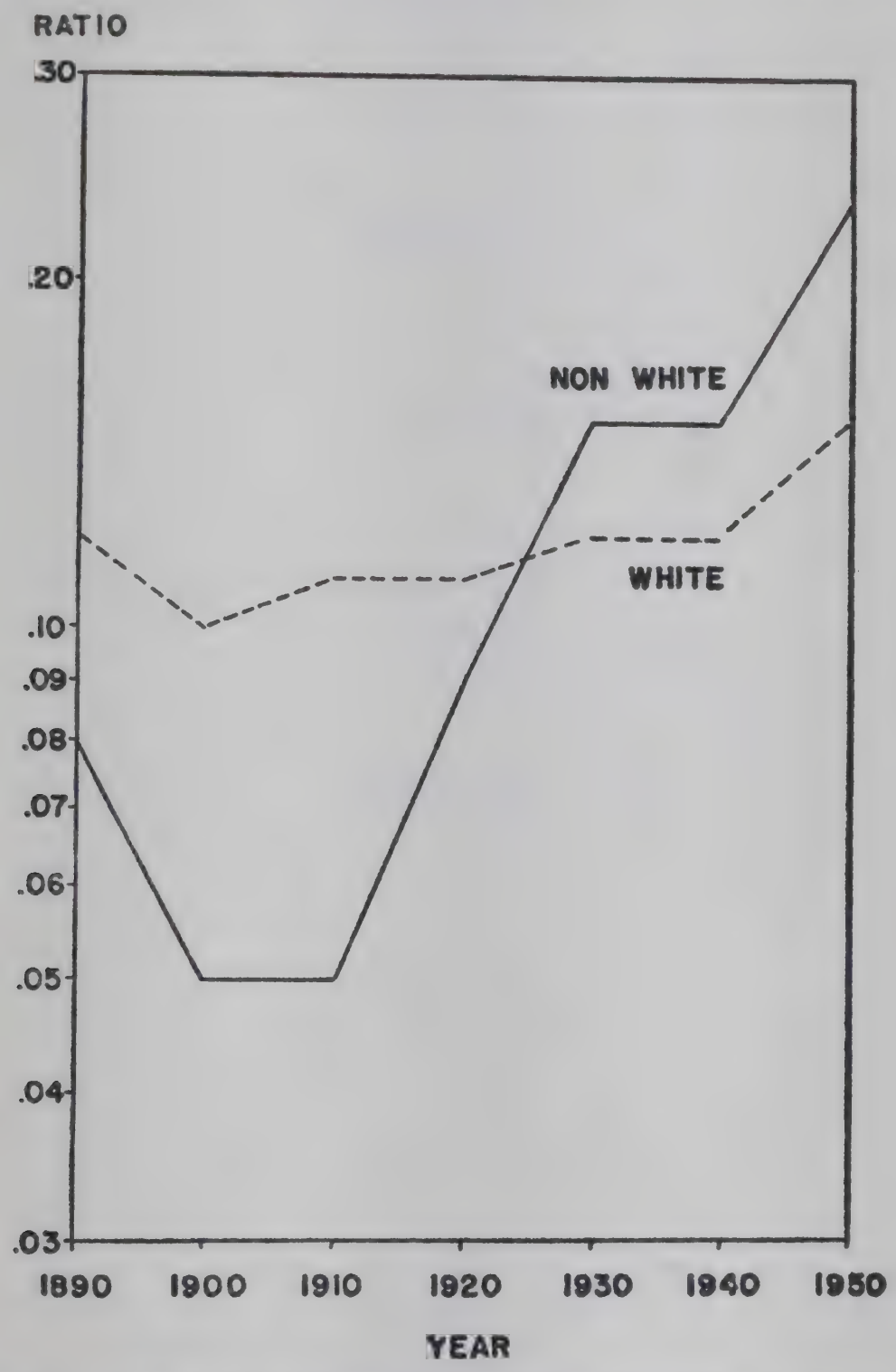


FIGURE 25. Ratios of the number of persons living in different regions from their region of birth to the number living in the region of birth, United States, nativeborn population, 1890 to 1950. (Data from U. S. Bureau of the Census, 1949, 1953a.)

TABLE 25

First admission rates for psychoses among "five-year migrants" and "nonmigrants," New York State, 1939-1941†*

Color and sex	Manic depressive		Schizophrenia		Other psychoses		All psychoses	
	Migrant‡	Nonmigrant‡	Migrant	Nonmigrant	Migrant	Nonmigrant	Migrant	Nonmigrant
White male	17	6	86	37	181	94	269	130
White female	37	14	76	35	133	74	231	116
Non-white male	10	6	261	74	804	267	1032	333
Non-white female	56	9	187	70	595	172	810	240

* Rates are average annual rates per 100,000 population and are standardized to the age distribution of the New York State population aged 14 and over in 1940.

†From Malzberg and Lee (1956).

‡ "Migrants" in these data are persons living in New York State on April 1, 1940 who were not living in New York State on April 1, 1935, while "nonmigrants" are persons resident in New York State on both dates.

movement. A question in the 1940 census asked "Where were you 5 years ago (April 1, 1935)?" This denominator enabled Malzberg and Lee (1956) to derive rates of psychosis among persons known to have made a recent (5 year) move for comparison with rates among those not known to have made a recent move. The findings are summarized in Table 25.

11

Time

The omnipresence of time renders impractical any extended attempt to consider it as a separable item of experience or reality. Time has been noted as an ingredient necessary to each of the various epidemiologic measurements and to the very concept of causal association. However, in a more limited context, consideration is given here to the frequent use of time measured in certain arbitrary units as a general localizing variable.

TIME IN VARIOUS LENGTHS

Time may be thought of in whatever size we care to — an instant or an eon. When marked changes in disease frequency are noted over a period of time judged very short in the comparative experience of the observer, explanatory hypotheses seem easier to propose (and are more readily accepted) than when the changes evolve more slowly.

"Point" Epidemics

Large and fleeting excesses in disease frequency are commonly called "point" epidemics. The same episode of cholera for which localization by place was shown in Figure 24 (page 154) is now illustrated in Figure 26 according to its occupancy in time. The localization by time is seen also to have been very marked.

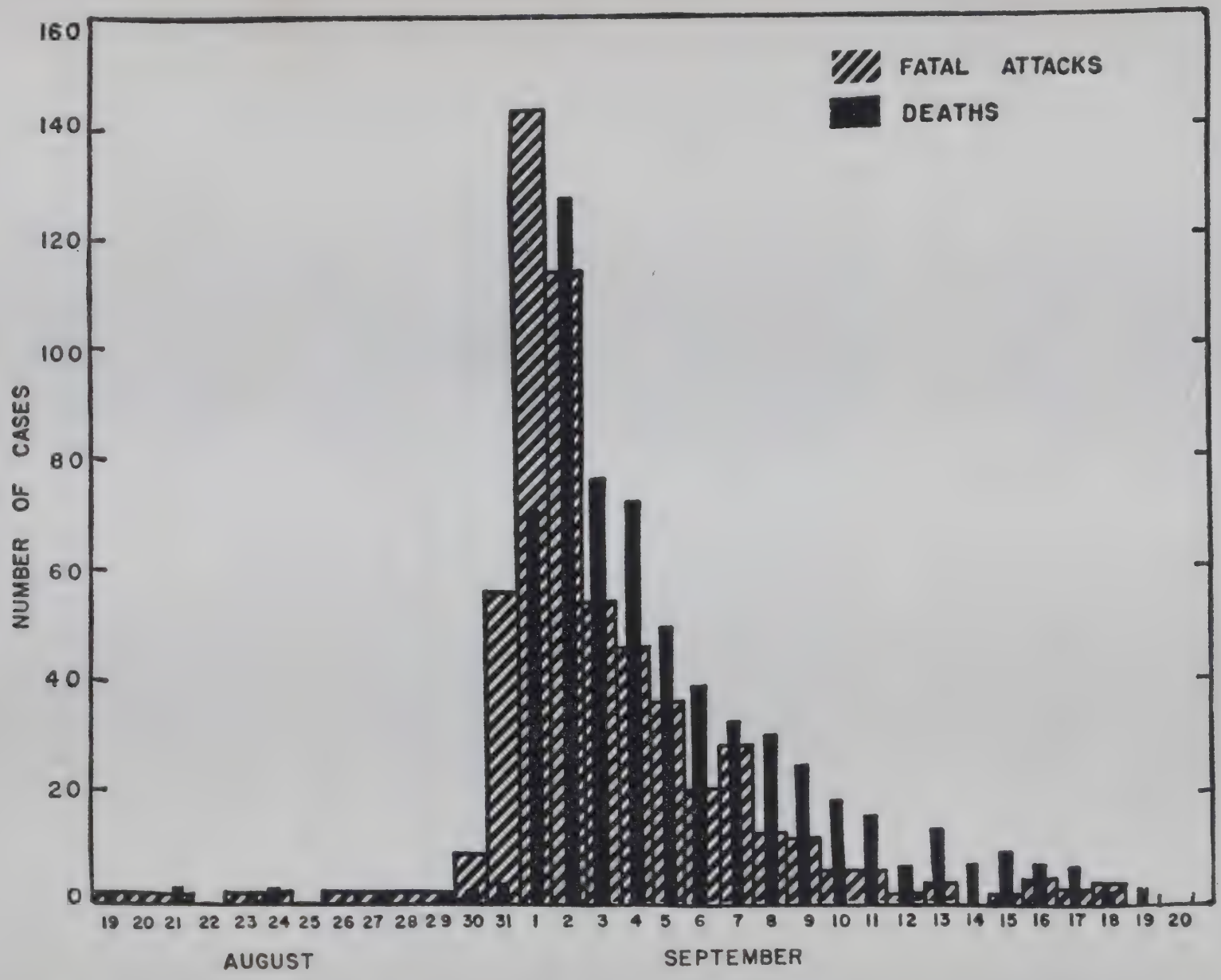


FIGURE 26. The outbreak of cholera in the Golden Square area of London, August and September, 1848. (Data from Snow, 1855.)

Such outbreaks of disease are particularly helpful in revealing etiologic factors, since the pattern (place and time) suggests that a large number of people were exposed to a similar causal circumstance almost simultaneously, and that the mean interval between exposure and onset was short. The latter supposition is based on the knowledge that some dispersion around the mean induction or incubation period between exposure to disease-inciting events and onset of disease is always present. For such dispersion to be encompassed within a few days indicates that the mean must itself be quite short. Both suppositions make search for the causal circumstance less difficult.

Marked rises in disease frequency occupying brief lengths of time have regularly been found associated with very clearly delineated circumstances such as abrupt encounters (via air, food or water) with microbiologic organisms or chemical agents. The rise and fall of epidemics is necessarily slower when engendered by more gradually introduced or less circumscribed forces. If lengths of time are conceived of in larger units, however, fluctuations in disease frequency that appeared gradual when smaller time units were used now present an appearance in time no different from that of a "point" epidemic. The dearth of serious attempts at identifying such epidemics in chronic disease is due probably in part to failure to realize that time actually is not delivered in predetermined lengths, and in part to the changing social scene that ensures that data with consistent definitions are seldom available over prolonged periods.

Secular Change

Secular change refers to changes that occur gradually over "long" periods of time, and in epidemiology usually implies changes in disease frequency measured over a period of several years or decades.

Morbidity data adequate to measure such changes have to date been limited to certain geographic areas and have been restricted to notifications of occurrences of infectious diseases and of use of mental hospitals. Although valuable in the detection of truly gross and rapidly evolving changes, clinical impressions are, of course, not dependable in detecting more subtle changes. Conse-

quently, death-certificate data have so far been a major source of information.

As measured by death certificate data, no disease has retained a constant frequency during the past 50 years. The investigation of such secular changes may be directed to two distinct questions:

(1) Has there been a change in frequency of the disease that is really of epidemiologic interest, or is the change due to variation in ascertainment, reporting, case fatality, or some other complicating circumstance?

(2) If the trend is of real epidemiologic interest, to what mechanisms may it be attributed?

Changing ascertainment encompasses a number of distinct possibilities. Changes due to altered procedure in assigning causes of death have already been discussed (see Chapter 6); they are usually readily identifiable because of their abruptness and coincidence with known procedural alteration. Change in case fatality rates is an obvious factor that has influence on secular trends in death rates. Inconstancy of the characteristics of the population must, of course, also be considered. Change in age distribution is an obvious and readily identifiable source of secular change; the independence of secular change and changes in age distribution can be ascertained by investigation of age-standardized or age-specific rates. Other changes such as altering proportions of various ethnic groups may not be so readily identifiable, or if identified or suspected, may not be so readily investigated because of absence of population or other data necessary for the computation of group-specific rates.

The component of the over-all question of epidemiologic relevance that is the most vexing, because it is the least readily answered, relates to the role of changes in clinical diagnosis, concepts, and terminology. This problem is most difficult to attack because such changes evolve gradually and thus do not have the telltale abruptness of procedural changes; further, while it is frequently possible to state with confidence that diagnosis of a particular condition is "better" now than it was 20 years ago, it rarely is possible to estimate how much better and to what extent the observed change in disease frequency may be attributed to the improvement. Any attempt to estimate how many cases of a

particular disease diagnosed at the present time would have been diagnosed with the knowledge and facilities available at some time in the past is usually unrewarding.

Assistance, however, in forming more than a subjective opinion on the role of changes in diagnostic capacity as an influence on disease trends may be derived from a number of procedures:

(1) Comparison of trends in the disease under investigation with trends in other diseases judged equally susceptible to added diagnostic facility. For example, in Figure 27, death rates from lung cancer in England and Wales during the period 1911 to 1955 are compared with death rates from cancers of several other internal sites. While lung cancer is by no means the only site that shows an apparent increase, the data leave no doubt that something has happened for lung cancer that has not happened for cancers of other internal sites, and if of a diagnostic nature, betrays a past ignorance that was remarkably localized from an anatomic point of view. Such evidence can only be suggestive because such a localized oversight is not inconceivable, and also because it is rarely possible to find two disorders in which diagnostic procedures have advanced *pari passu*.

(2) Comparison of the frequency with which the disorder is present at autopsies performed at different times during the observed secular change. Paradoxically, this procedure will be increasingly satisfactory the less fatal the disease under investigation; that is, the greater the probability that the patients died of some other cause. The number of patients with the disease under study that is observed will, then, be a measure of the prevalence of that disease, and may be related to the number of autopsies performed to derive a point prevalence rate (see Chapter 5, page 57). In a fatal disease, however, the measure will be more directly related to incidence, but it will probably not be possible to convert the autopsy data to rates because of inability to define the population to which the deaths relate. The restriction of this method to institutions where a high proportion of deceased persons underwent autopsy examination in all compared periods reduces errors resulting from changing patterns in the selection of deaths for autopsy. However, it does not eliminate the possible effect of changes in the types of patients entering the institution before death.

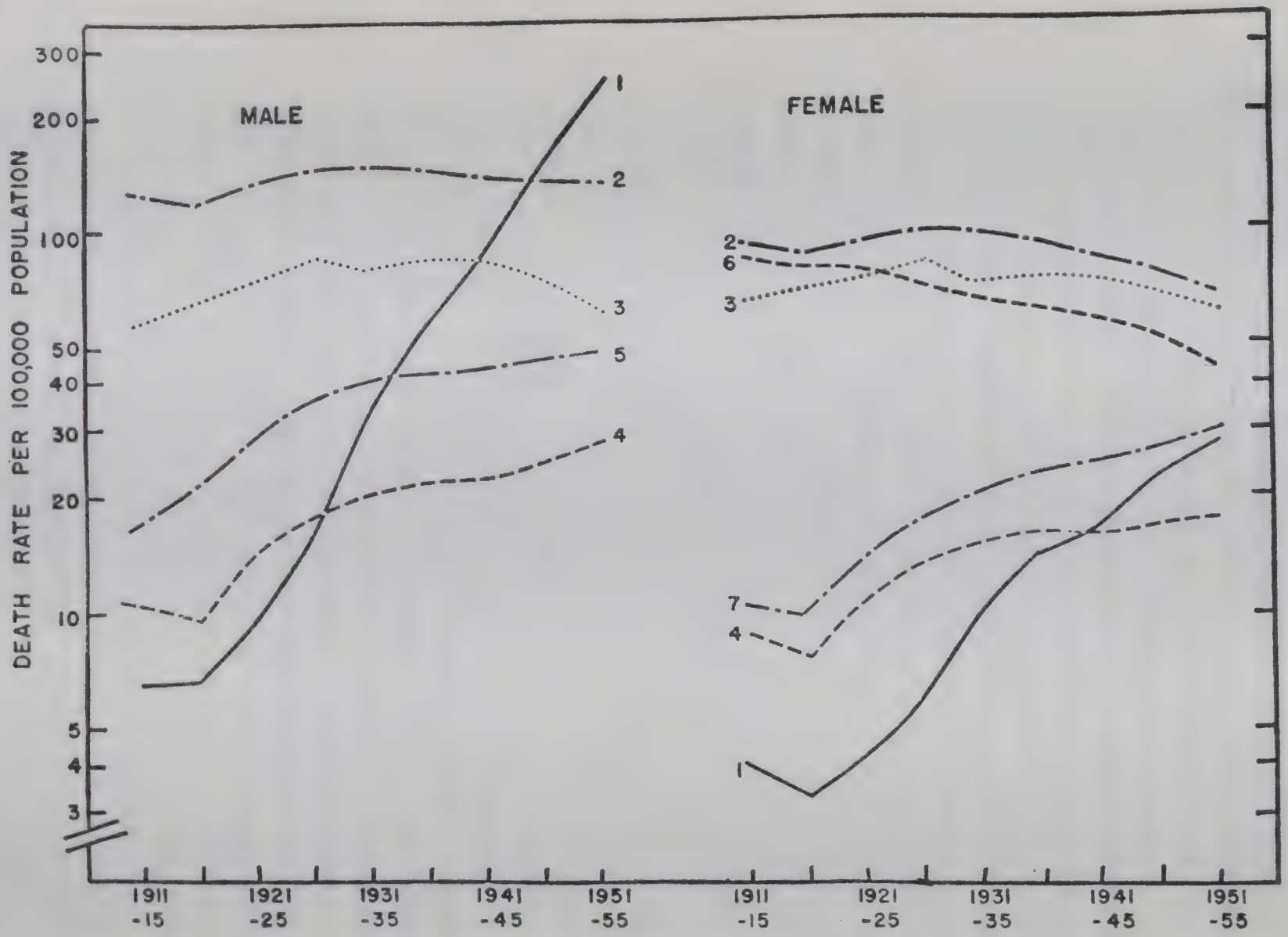


FIGURE 27. Equivalent average annual death rates from selected sites of cancer in persons aged 50 to 74, England and Wales, 1911-1955. Rates shown are the averages of those for the five 5-year-age intervals between 50 and 74 given by Case and Pearson (see legend for Figure 18). Sites shown are: (1) lung, (2) stomach, (3) intestine, excluding rectum, (4) pancreas, (5) prostate, (6) uterus, (7) ovary. (Data from Case and Pearson, 1956.)

(3) Examination of the extent to which the apparent trend might be influenced by secular change in the proportion of cases misdiagnosed, as based on a variety of assumptions, with consideration of the reasonableness of the assumptions necessary to eliminate the trend. This method, although used infrequently, is illustrated by the examination by Gilliam (1955) of the possible effect on the trend in lung cancer death rates had various proportions of patients whose deaths in previous years were certified as due to other respiratory diseases in fact had lung cancer. Because the number of deaths attributed to other respiratory diseases was much larger than the number attributed to lung cancer, particularly in the earlier years of this century, a quite small proportion of the other respiratory disease deaths, if they had in actuality been lung cancer deaths, would have made a substantial addition to the number certified as due to lung cancer, and therefore a consequent substantial reduction in the apparent trend. However, the size of the assumed proportion of misdiagnosed cases necessary to eliminate the trend was so large that the assumption did not seem reasonable when considered in conjunction with such autopsy material as was available in the literature.

Lastly, there are a number of general arguments that can be applied to a discussion of the epidemiologic meaning of an apparent secular trend. Thus, trends that are markedly different for different age, sex or ethnic groups living in the same community would seem less likely to be artifacts than if they were not so localized within special subgroups within the community. Also it may be illuminating to ask, for example, in the case of lung cancer, whether diagnostic ability has become more sophisticated in recent years (say, between 1950 and 1959) at the same rate as it did during the earlier years when such diagnostic aids as X-ray and bronchoscopy were being developed. In lung cancer the upward trend has shown no evidence of recent leveling off.

While there can be no doubt that lung cancer has undergone a real and substantial increase during the last three decades, it must be noted that secular trends of this magnitude are unusual in the noninfectious diseases. Much more frequently, as in the apparent increases in death rates from coronary heart disease, leukemia, and pancreatic and ovarian cancer, and in the increase in the trend of patients admitted to mental hospitals, the issue whether

an observed secular change reflects a change in disease frequency remains in doubt in spite of use of all the procedures and arguments mentioned. Although this uncertainty regarding the actual mechanism of the secular change must be given consideration in the formulation of hypotheses of etiology, outright denial of the possibility that the change may be of real etiologic significance is illogical.

Interpretations

The kind and magnitude of social, biological, physical and chemical circumstances that change with the passage of time are as infinite as time itself. An attempt at their enumeration would be purposeless. Discussion of general circumstances related to the occurrence of "place" diseases (see Chapter 10) is equally relevant in a consideration of the role of time, but is not repeated here.

We might note, however, that certain interpretations are more easily eliminated from consideration in place associations than in time associations. These are the ones discussed under changing ascertainment (page 142), since the conditions of various periods of time cannot be subjected to the special surveys that can be conducted to test a hypothesis of association with place. On the other hand, interpretations involving genetic differences in populations are usually more easily eliminated as productive of time associations. Associations with time occurring within relatively short periods such as decades can only be attributed to genetic changes if there have been marked population shifts. And such shifts will usually be identifiable.

It has been repeatedly noted that changes occurring over a relatively short time may be more productive of hypotheses than those occurring gradually over decades. This is true even when the changes do not have the abruptness of a "point" epidemic. An example, given in Chapter 4, is Snow's observation of a change in the incidence of cholera in the customers of the Lambeth Company over a 5 year period. The importance of the observation of such a relatively rapid change lies in the restriction of the period that must be examined for changing conditions possibly productive of the observed change in disease frequency. A more recent example is illustrated in Figure 28, which shows a striking and progressive increase between 1941 and 1945 in the

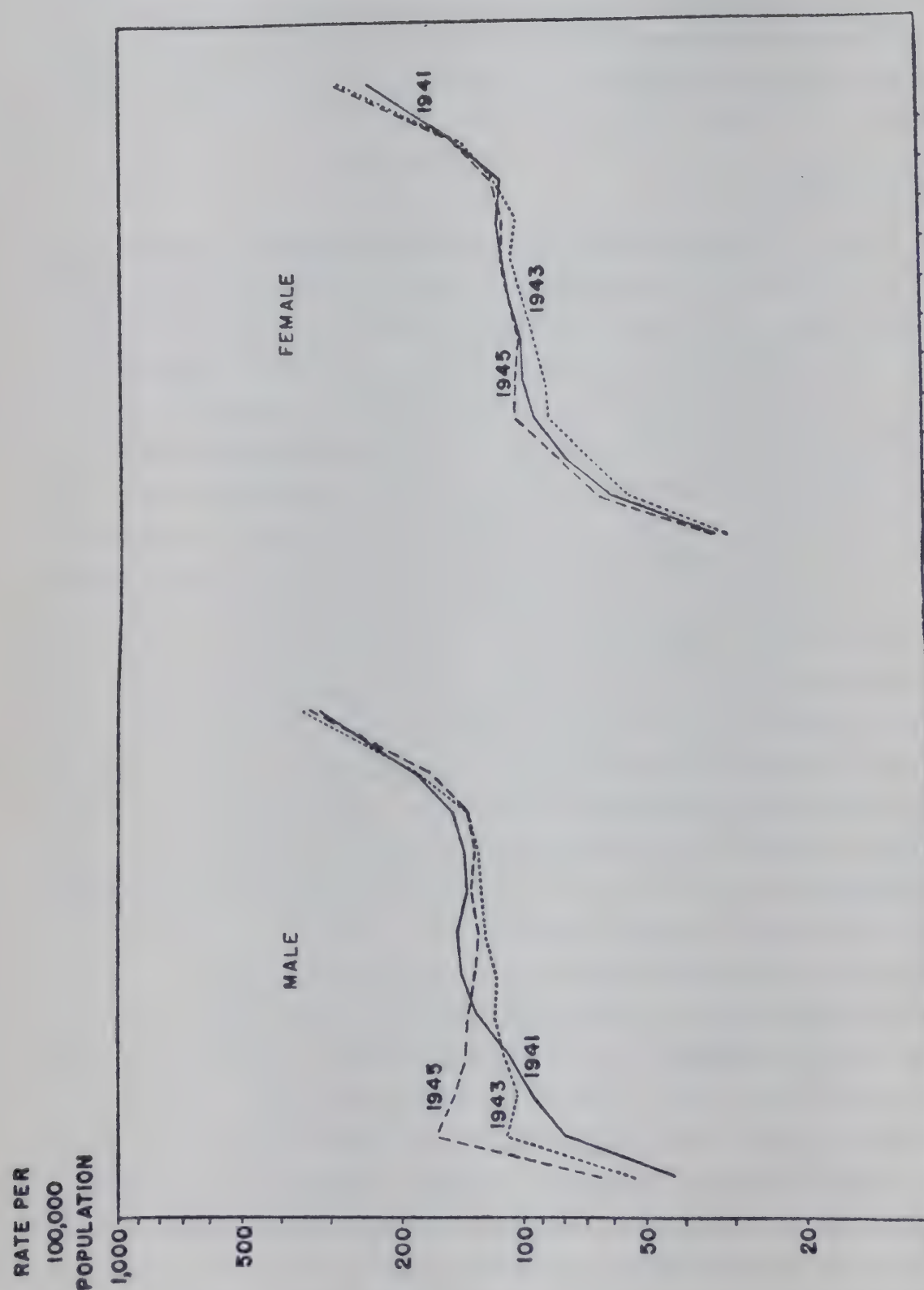


FIGURE 28. Incidence of first admission to mental hospitals per 100,000 population by age and sex, United States, 1941, 1943, and 1945. (Data from U. S. Bureau of the Census, 1944, 1946, 1948.)

rates of first admission to mental hospitals in the United States for males between the ages of 15 and 34, the greatest increase being for the age group 20 to 24. The pattern suggests that the stresses of military life may have an etiologic influence in mental disorder, at least among persons who under ordinary circumstances would not have elected to enter the service. An alternative hypothesis — that induction into the service acted as a case-finding mechanism for this age and sex group, and hence, led to increased hospitalization rates — no doubt plays some part, but is unlikely to be a complete explanation since the trend persists until 1945, while the number of persons being inducted and examined declined markedly after 1943.

Cohort Analysis of Secular Change

The principle of cohort analysis as described in the discussion of association with age (see Chapter 7) also has application to the study of time trends. Thus, in the conventional method of showing trends in age-specific rates over time, as illustrated in Figure 29, the abscissa of the graph is the year of death or occurrence of the event being measured. Therefore, the age-specific rates shown at any point along the abscissa are based on persons belonging to different cohorts. We see in Figure 29 that, in the oldest age groups, death rates for buccal cancer increased during the early years of the period and decreased in the later years, producing a peak in the trend about the middle of the period. There is a tendency for this peak to appear successively earlier in time as successively younger age groups are examined, until, among persons younger than 50, rates show a downward trend during the whole period of observation.

The question might be asked whether this pattern can be explained by the peak rates being characteristic of a particular generation of males and whether the later appearance of the peak as age advances can be explained by this generation occupying successively older age groups in successively later periods of time. In other words, are the secular changes noted for buccal cancer characteristic of the experience of cohorts, rather than a result of differential changes in the environment of persons of different age with the passage of time? This question may be examined by plotting the data with year of birth, rather than year of occur-

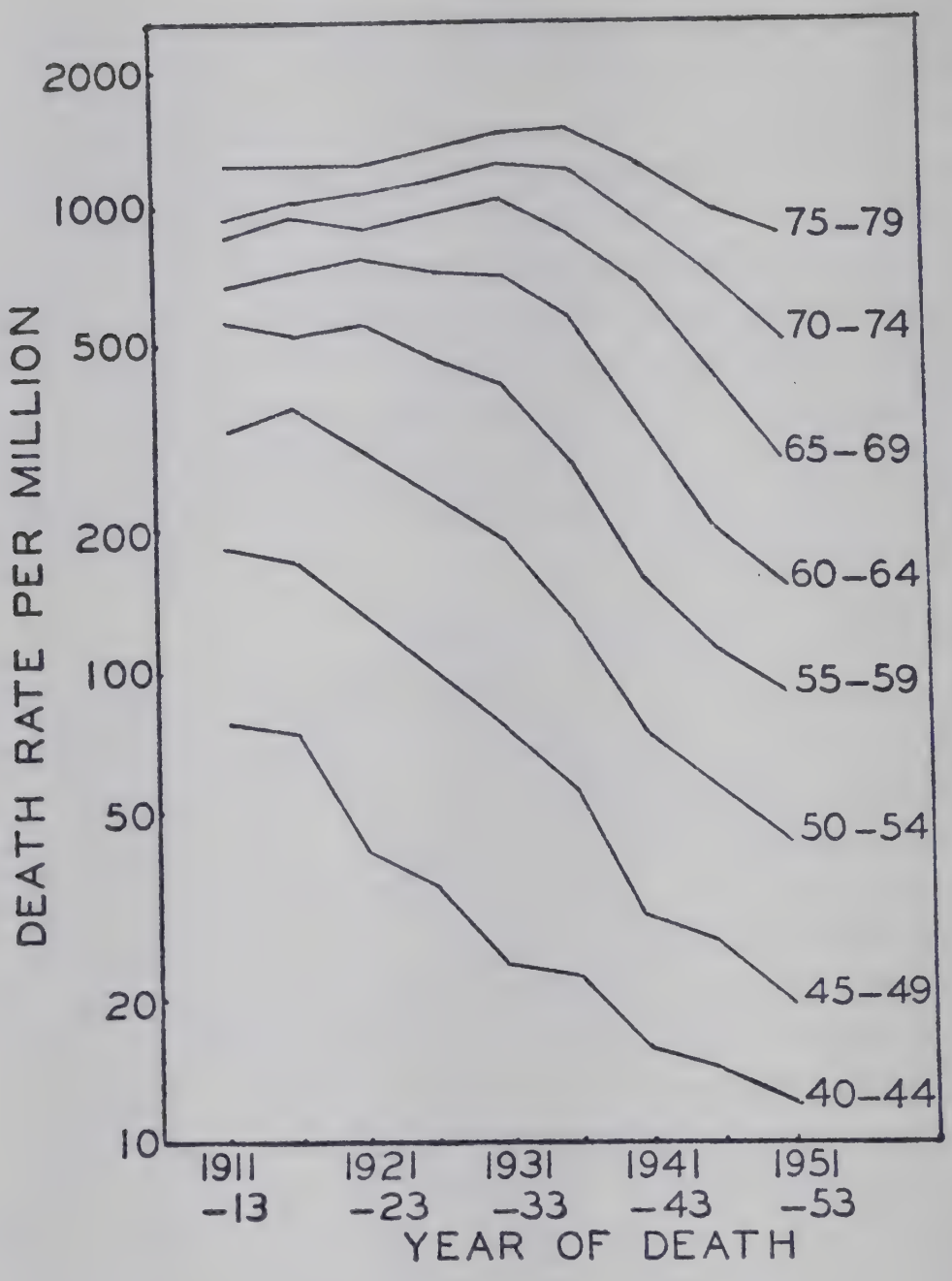


FIGURE 29. Age-specific mean annual death rates from cancer of the buccal cavity and pharynx, England and Wales, 1911-1954. (From MacMahon and Terry, 1958.)

rence, along the abscissa, as in Figure 30. A trend noted between any two points on the abscissa now reveals differences between successive cohorts rather than between successive time periods. We now see that, in contrast to Figure 29, the patterns exhibited by the various age groups are similar. At all ages, death rates increased with each successive cohort and reached a peak in persons born around 1865; after 1865 each successive cohort born has experienced lower death rates from buccal cancer, at all ages, than did its predecessor.

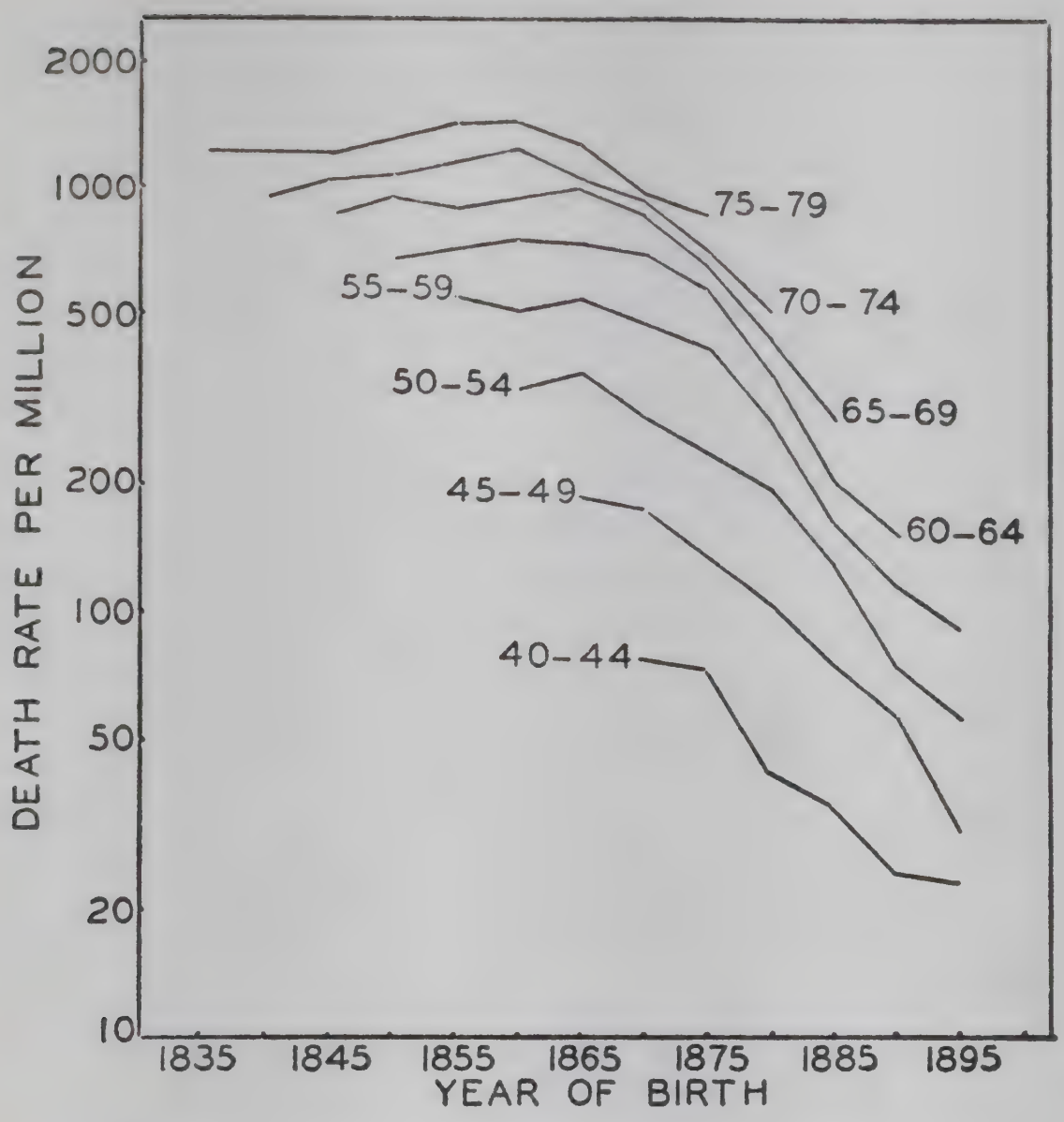


FIGURE 30. Data from Figure 29 arranged to show date of birth of the populations for which rates are expressed.

As noted in Chapter 7, the decision whether a cross-sectional or a cohort approach is appropriate to a particular situation depends on (1) a priori reasoning based on knowledge of the natural history of the disease in question, and (2) evidence of a greater consistency found by one or the other method. It was also noted that many diseases, such as the degenerative diseases, do not show any peak in their age curves that allows examination for a greater consistency by one or the other method. The method illustrated in Figure 29 and Figure 30 offers another possibility in such situations, provided that during the period of observation rates for the disease examined have undergone a change in direction so that a peak (or a trough) in the time series is produced. It can

then be determined whether the peak, or the trough, occurred consistently (that is, at all ages) in a particular period or in a particular cohort. Figure 30 suggests that for buccal cancer the changes are characteristic of cohorts, and patterns in a number of other neoplastic disorders are also compatible with that hypothesis (MacMahon and Terry, 1957). Several other neoplastic disorders and other degenerative diseases do not allow application of this method because trends have been consistently in one direction during the period for which frequency data have been available.

The general implications of the observed patterns of change that are characteristic of cohorts so far as hypotheses of etiology are concerned have been discussed in Chapter 7. There are two other considerations that are particularly relevant in connection with trends in age-specific rates over time. First, changes occurring in a cohort pattern are unlikely to be explicable in terms of changes in effectiveness of therapy, diagnosis or reporting, and must be sought in changes in etiologic factors. It is of course possible that, for example, effectiveness of therapy might vary with age and produce trends that were not consistent in different age groups. But it would be a strange coincidence if the change in effectiveness of therapy became applicable to successively older 5 year age groups in successive 5 year periods, as would be necessary to explain a cohort pattern.

The second consideration relates to methods of prediction of time trends.

Prediction of Time Trends

It is sometimes useful for administrative purposes to be able to extrapolate time trends beyond the present available data. In this context the distinction between current and cohort patterns is important.

For example, age-specific death rates from breast cancer in females in England and Wales between 1915 and 1954 are shown in Table 26. The crude total rate (all ages) shows an increase in death rate over time. However, this must be largely attributed to a change in the age composition of the population, since in most age groups the trend has been downward since the period 1935-39 and the total age-standardized rate decreases consistently

TABLE 26

*Mortality rates for breast cancer per 100,000 female population, England and Wales, 1915-1954**

Age at death	Year of death							
	1915-1919	1920-1924	1925-1929	1930-1934	1935-1939	1940-1944	1945-1949	1950-1954
Under 25	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25-29	0.8	0.8	1.2	0.8	1.0	1.3	0.9	1.2
30-34	3.7	4.0	4.6	4.5	4.8	4.7	5.4	5.2
35-39	12.6	12.1	12.7	13.3	12.2	13.5	14.1	14.2
40-44	26.6	27.0	27.7	27.7	26.9	26.9	28.2	29.3
45-49	47.6	47.1	46.5	47.5	45.5	45.6	44.4	47.1
50-54	61.3	60.6	63.7	65.7	63.8	61.3	58.6	56.8
55-59	73.5	77.9	74.9	80.2	80.4	75.0	73.1	70.9
60-64	84.0	88.4	92.3	97.4	97.4	90.5	86.4	84.1
65-69	97.1	98.4	108.4	106.8	111.0	106.4	100.9	99.9
70-74	117.9	122.7	131.2	131.4	134.1	118.3	120.8	116.6
75-79	151.4	160.0	166.0	168.4	174.1	138.9	148.5	143.9
80 or more	203.1	225.5	228.9	244.9	241.4	191.5	193.6	195.8
Total, crude	21.1	24.2	27.5	30.7	33.2	32.9	34.7	35.7
Total, age-standardized†	27.6	28.5	29.4	30.3	30.2	28.1	27.7	27.5

* Data from Registrar General's Statistical Review.

† Standardized to the age distribution of the female population in 1931.

Time

since the years 1930-34. On this basis, therefore, we might predict that age-standardized death rates from breast cancer might continue to decline.

However, examination of trends in the individual age groups reveals at least two distinct patterns:

(1) A peak occurring in the period 1935-1939 and a subsequent decline are most striking in the older age groups and are not evident in women younger than 50 years. This peak is evidently characteristic of current time, rather than of a cohort, since it is seen in four of the six age groups over 50, all of which belong to different cohorts. (In the other two age groups over 50 the pattern is similar but the peak occurs in the years 1930-1934.)

(2) A tendency to increasing rates is present in the younger age groups during the latest periods. This upward swing does not occur in the same period in each age group, and is in fact suggestive of a cohort pattern. Thus the first upward trend occurred in the period 1950-54 in the age group 45-49, five years earlier in the age group 40-44 and five years earlier still in the age group 35-39. The pattern suggests that, at all ages attained to date, women born around 1900 had the lowest breast cancer death rates, and that rates have increased for successive cohorts born thereafter.

On the basis of this latter pattern we might therefore predict that as women born after 1900 reach those age groups in which breast cancer rates are highest, breast cancer death rates will in fact increase. This prediction is, of course, based on the assumption of a continuation in the cohort pattern. The assumption would be made meaningless by the subsequent occurrence of, for example, a marked improvement in therapeutic efficacy, which would be expected to be characteristic of a time period rather than of a cohort.

With respect to the forecasting of trends in general death rates, there has been considerable controversy among demographers as to whether cohort or current methods should be used (Pollard, 1949). The problem seems to us an empiric rather than theoretical one, and can only be decided by examination of past data to determine whether cohort or current patterns are occurring. Actually, in the context of so complicated a mechanism as the determinants of total death rates it is quite certain that mech-

anisms following both types of patterns are operative. In the prediction of trends in individual diseases more simple patterns may be evident and form the basis for predictions which will come closer to the mark.

CYCLIC FLUCTUATIONS

If this text were concerned with the content of epidemiologic knowledge rather than with epidemiologic methods, it would be necessary to devote considerable space to periodic variations in disease frequency. For example, seasonal fluctuations are among the most striking features in the epidemiology of the infectious diseases and have received a great deal of attention.

However, the methods involved are comparatively simple. Cases are plotted by time of onset according to a conventional time unit. In detection of seasonal fluctuations, months or quarter years are generally used. In order to detect periodicity of any type it is necessary to use time units which are some fixed fraction of the interval between two successive fluctuations. Use of a reference population and the calculation of a series of rates are unnecessary refinements, unless periodic changes in the size of the population are suspected.

Unlike the acute diseases, chronic diseases have received little attention with respect to possible periodicity of occurrence. This is probably the result both of difficulty in defining date of onset in chronic diseases and of belief that etiologic factors in chronic diseases are operative through such prolonged induction periods that correlation with some other periodic variation (such as season) would be unlikely.

Interpretations of Seasonal Variations

Methods of utilizing knowledge of seasonal periodicity to formulate hypotheses center around the process of correlating seasonal variation in disease frequency with seasonal variation in other circumstances of possible significance. For example, many studies have related disease occurrence to weather conditions (and hence to season). Figure 31 shows monthly deaths and monthly normal temperatures for three cities in the United States. As noted by many writers, the inverse correlation is strik-

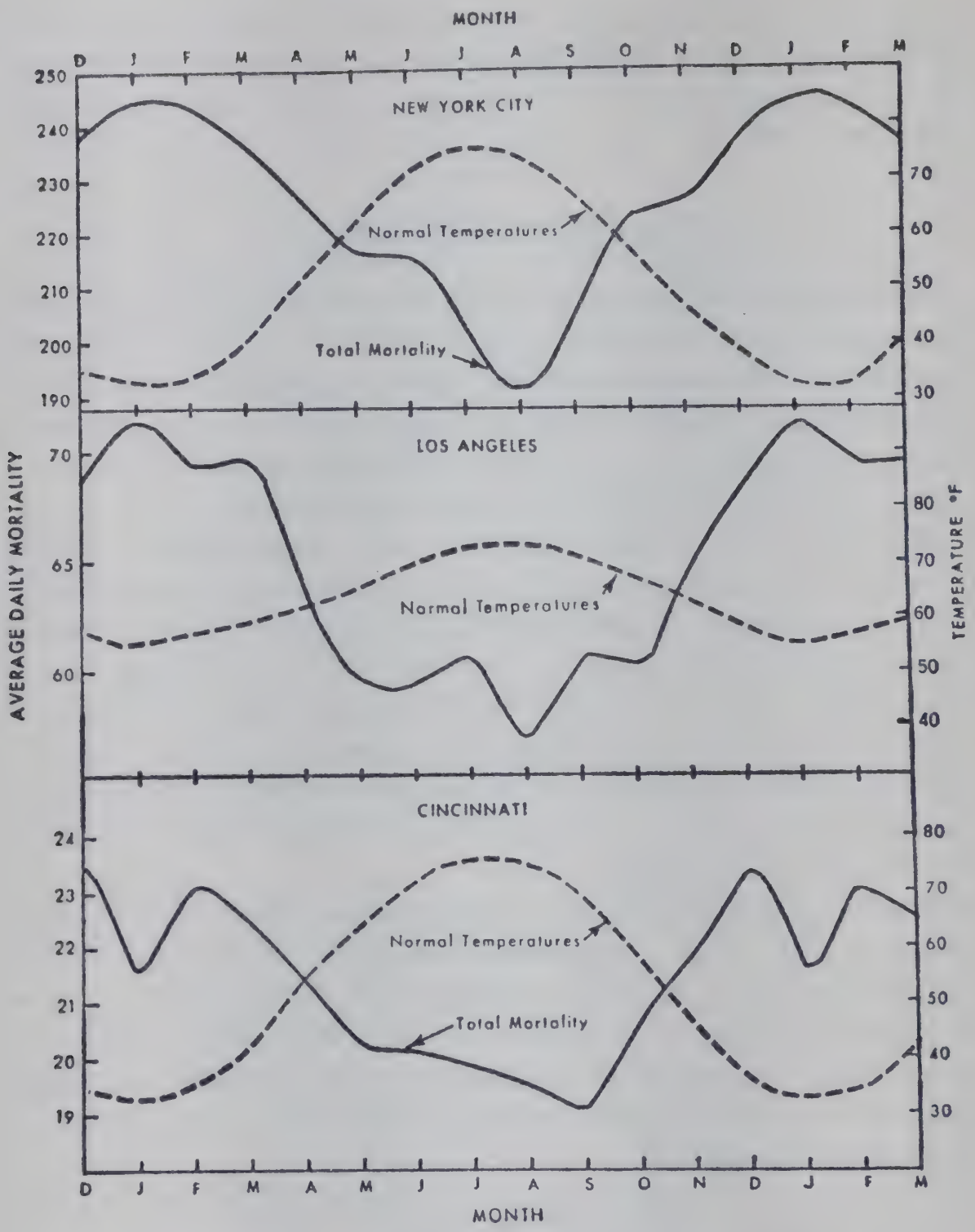


FIGURE 31. Seasonal variation in mortality and temperature in three cities. The temperature curves are based on monthly normal temperatures for 1921-1950; mortality curves are based on monthly totals of deaths, 1949-1958, computed on the basis of daily averages. (Reproduced, with permission, from Kutschenreuter, 1959.)

ing. However, it should be noted that although seasonal fluctuations in mortality are as great in Los Angeles as in New York City, the seasonal temperature fluctuations are smaller in Los Angeles than in New York City. Similarly, the temperature

fluctuations are greater but the mortality fluctuations less in Cincinnati than in Los Angeles. These inconsistencies suggest that, in spite of the inverse correlation, normal temperature variations account for only a small part, if any, of the seasonal mortality pattern. No greater success has been achieved in correlating other indices of seasonal climate variation with mortality, although it is likely that short periods of highly unusual climatic conditions, such as heat waves, have an appreciable effect on mortality.

Seasonal variation in the flora and fauna of the human environment has a profound effect on many human diseases and has been the area in which greatest use has been made of periodic variation in the development of hypotheses of causation. The resulting investigations have ranged from the identification of pollens concerned in hay fever to the search for insect vectors whose seasonal reproductive and feeding patterns are such as to explain seasonal patterns in disease frequency. For example, diseases contracted from ticks, such as Rocky Mountain spotted fever, show marked seasonal localization in the spring and early summer, corresponding to the period of feeding activity of the adult ticks. The mosquito-borne encephalitides, on the other hand, are localized in the late summer and early fall.

Seasonal fluctuations in human activities may also account for variation in exposure to sources of infection. Thus, the marked concentration of human leptospirosis during the summer months can be accounted for by increased exposure to infected waters through bathing and fishing. The crowding of persons together during the cold weather leads to conditions conducive to the spread of epidemic typhus.

Seasonal variation in the acute infectious diseases of childhood constitutes one of the major unsolved problems of infectious disease epidemiology. In spite of the consistency and striking magnitude of this variation, it remains unexplained; yet if, for example, the factors responsible for the seasonal variation in poliomyelitis were known and lent themselves to practical alteration for preventive purposes, the disease would cease to be a problem. Simpson (1952) showed that there was no seasonal variation in the secondary attack rates of measles, chickenpox and mumps once an infection had been introduced into a household. Seasonal variation in virus communicability does not therefore seem a

likely explanation of the seasonal variation in these diseases.

Incidentally, seasonal circumstances that are entirely man made may play a part in seasonal disease patterns. This is illustrated by the study of Friedman *et al.* (1958) which showed rises in serum cholesterol levels of accountants corresponding to critical dates in the tax calendar. If this is a real effect, it is unlikely that it is limited to accountants.

For diseases in which environmental influences acting during intrauterine or early postnatal life may be of etiologic significance, season of birth is an informative descriptive variable which may point to the existence of such influences and give hints as to their nature. Probably the most clearly substantiated relationship of this type is that between intelligence and season of birth. Children born in the winter months have a significantly lower mean score on intelligence tests than those born in the summer months (Pintner and Forlano, 1943). In addition, risk of admission to a state school for the mentally retarded was shown by Knobloch and Pasamanick (1958) to be highest for children born in February and lowest for those born in August. A number of hypotheses have been developed as a result of this observation, one of which relates the mental retardation to unidentified nutritional disturbances in early pregnancy, these disturbances being caused by changes in maternal diet during the hot summer months. In an interesting investigation of this hypothesis, Knobloch and Pasamanick compared admission rates for mental retardation for children whose critical stage of intrauterine development (in this case, supposed to be the third month) occurred during summers in which the temperature was high with those born following summers in which the temperature was low. Admission rates were significantly higher for the former group than for the latter. This is an interesting combination of season and secular variation which deserves further elucidation.

CLUSTERING IN TIME

The descriptive studies discussed so far have all been based on the comparison of rates that involve knowledge of reference populations. As noted in the discussion of spot maps, "point" epidemics and periodicity, comparisons are sometimes made on the basis

of numbers of cases, but knowledge of the population is implicit. Thus, in the case of an epidemic such as that of the London fog, it is assumed that the population base remained the same throughout; thus, the principle of comparison of two or more populations with respect to disease frequency still holds.

However, it occasionally happens that hypotheses of etiology are derived from the clustering of certain events along the time axis in the histories of affected individuals, even though no inferences regarding the population at risk are made and no comparisons with other groups are involved. This time clustering is best seen in diseases in which the interval between precipitating event and onset of illness can be measured with precision.

For example, Aycock and Luther (1929) obtained histories of tonsillectomy from a series of patients with poliomyelitis. They noted that among 36 cases of poliomyelitis that occurred within 12 months of tonsillectomy there were 16 in which poliomyelitis followed within 1 month of the operation. All these 16 cases occurred within 7 to 18 days of tonsillectomy, a period the authors considered compatible with the incubation period of poliomyelitis. Even though no data were available to determine how many children selected at random might have given a history of tonsillectomy within 7 to 18 days, the marked clustering of tonsillectomies strongly suggests a causal relationship between tonsillectomy and poliomyelitis in at least some of the cases.

Similarly, data of Tetlow (1955) on the distribution by month of 67 patients in whom mental disorder was associated with pregnancy or the puerperium show a marked clustering of cases with onset of symptoms in the 4 weeks after delivery (Figure 32). Within the group with onset in the 4 weeks after delivery there was additional clustering, nearly half of them becoming ill in the week immediately following delivery. Such a clustering in time indicates the probable presence of psychotogenic factors operative around the time of delivery of a child at least among some of the women observed.

A more statistically sophisticated variation on this theme is the measurement of the interval between the suspected event and onset of the disease in a series of cases, and the determination of the mean and variance of this interval. The variance of the interval should be smaller if the measurement is taken from the date of

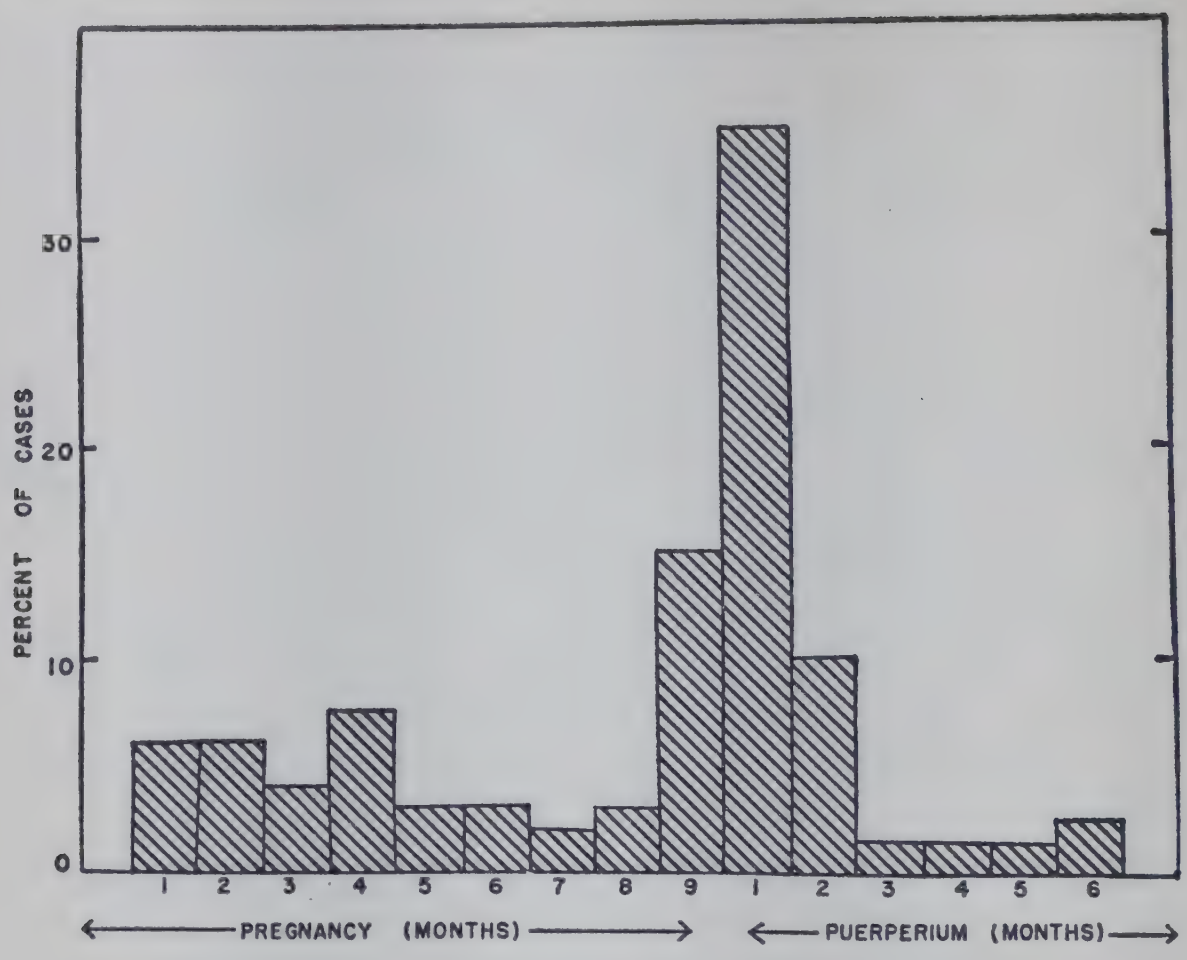


FIGURE 32. Distribution by time of onset of symptoms of 67 patients with mental disorder associated with pregnancy and the puerperium. (From Tetlow, 1955.)

occurrence of a causally related event than a variance derived from a mean of points in time that have a less direct causal connection to the disease. This method was used in an investigation of the possible relationship of hepatitis among service personnel to yellow fever vaccination (Sawyer *et al.*, 1944). It was impossible to compare affected and unaffected individuals with respect to frequency of yellow fever vaccination, since practically all the military personnel under observation had been vaccinated. However, investigation of the time relationship showed that the interval between vaccination and onset of the disease was less variable than the date of onset. The mean date of onset for the 5917 California cases was May 31.9, and its standard deviation 5.7 weeks, while the mean interval between vaccination and onset of jaundice was 14.6 weeks and its standard deviation 3.5 weeks (Figure 33).

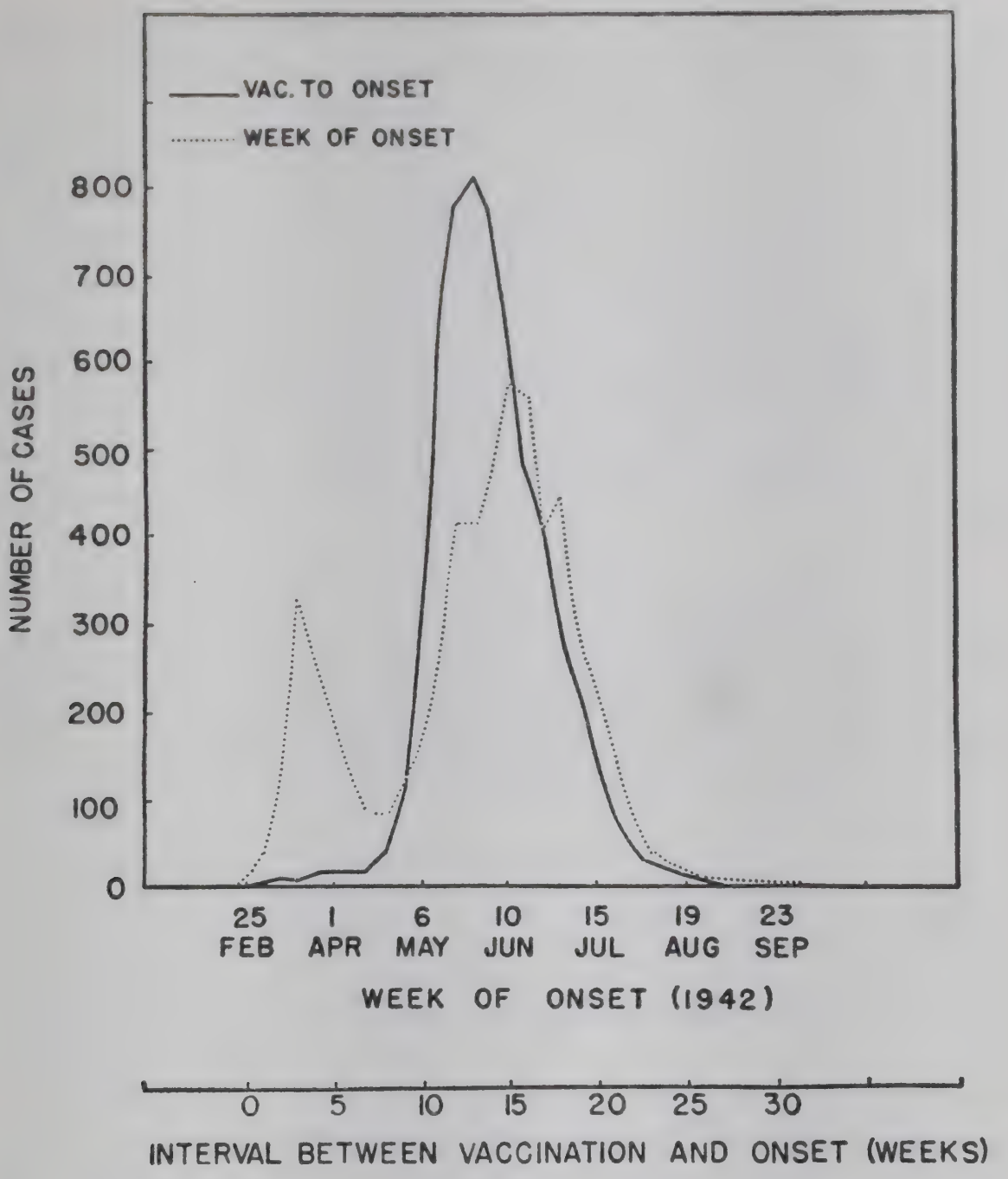


FIGURE 33. Distribution according to week of onset and according to interval between yellow fever vaccination and onset of the disease for 5917 cases of jaundice, U.S. military units in California, 1942. (From Sawyer *et al.*, 1944.)

With respect to clustering of events along the time axis in the histories of affected individuals, it should be noted that:

(1) For clustering to be observed with greatest clarity the interval between effect and presumed cause should be measured with as much precision as possible, in order to keep as small as possible that part of the variance derived from errors in measurement.

(2) Clustering will be less apparent when the interval between effect and cause is very variable, and additionally so if there is a consistent change in this interval during the course of an epidemic. For example, if the interval between exposure and onset decreases as an epidemic progresses (as may be the case in an outbreak of staphylococcal food poisoning), it is possible that the actual dates or times of onset of cases may be more closely grouped about their mean than the intervals between exposure and onset are about theirs. Consequently, the fact that the variance of the interval dated from a certain event is not smaller than the variance of, for example, the date of onset does not necessarily indicate a lack of causal connection between the event and the disease.

(3) While clustering in time does indicate that a certain preceding event and the onset of a certain disease are associated in time, the same reservations about the causal nature of such an association must be made as are made about associations based on relative frequency of concordance of two events.

12

Birth Order and Parental Age

In the investigation of certain categories of disease or defect there is especial interest in the identification of environmental influences operative during intrauterine life. This is particularly true of disorders manifest at birth or in infancy, for example, the congenital malformations, although it is equally conceivable that prenatal environmental characteristics may be causally associated with mental and physical attributes and defects that are manifested much later in life.

Through astute clinical observation certain noxious agents operative during prenatal life have been brought to attention. For example, Gregg (1942) noted the occurrence of maternal rubella during a number of pregnancies productive of offspring who were congenitally blind from cataracts. Subsequent special studies confirmed Gregg's suspicions as to the causal nature of this association. In similar fashion some of the undesirable effects of prenatal irradiation, maternal syphilis and toxoplasmosis have been identified. The epidemiologic methods used in the testing of these clinical suspicions are described in subsequent chapters.

The specific agents identified to date account for only a minute fraction of the cases of disease and defect in which it seems logical to believe that prenatal environmental influences may play a part. The vast majority of cases of congenital malformation and

mental retardation, for example, remain unexplained. In addition, there are large areas, for example, in the mental disorders, in which it seems reasonable to look for evidence of prenatal influences even though the mechanisms by which such influences might be operative are not obvious. Conceivably other specific agents will be identified through direct clinical observation. Further, the animal-vegetable-or-mineral approach of descriptive epidemiology is appropriate to such problems; this approach will help to determine whether it is likely that prenatal environmental influences of some nature are causally related to particular manifestational disease categories and to localize investigative efforts within the potentially most profitable areas.

In addition to the general descriptive variables that have been discussed in previous chapters, there are certain other variables peculiarly appropriate to this area of investigation. The two that have the widest application are order of birth in the sibship and age of parents at the time of birth. These two variables are the subject of this chapter. It should perhaps be made clear that we are concerned here with the effects of parental age and birth order (or "birth rank") on characteristics of the offspring, and not with their effects on disorders of reproduction exhibited by the parents.

METHODS

Birth certificates, obstetric records of hospitals, and interviews are the common sources of information on the distribution of a series of patients according to birth order and maternal age.

There are two general methods of examination:

(1) Comparison of the birth order or parental age distribution of the series of affected individuals with the population of births from which they were drawn, or with a representative sample of that population.

(2) Comparison of the birth order or parental age distribution of the series of patients with the expected distribution calculated from the size and other characteristics of the sibships into which the patients were born.

Comparison with the related population, or with a sample of it, is the method of choice in studies in which the interval between the birth of the series of patients and the date of the study is

short. This requirement is most likely to be met in studies of characteristics which are diagnosed at birth or shortly thereafter. The method is difficult to apply in studies undertaken many years after the birth of the affected individuals, such as in studies of diseases first appearing in adult life, since in these circumstances it is usually difficult to identify and obtain information on the related population of births.

Fortunately, the second method, described by Greenwood and Yule (1914), is particularly applicable to studies in which there has been a long interval between the birth of the affected individuals and the time of the study. The basic assumption in the Greenwood-Yule method is that an affected sibling is equally likely to be born at any of the birth ranks or parental ages represented in the sibship to which he belongs. This assumption is valid only if the families under study are complete — that is, if the parents responsible for the sibship have completed their reproduction. The Greenwood-Yule method is therefore particularly suitable for the examination of conditions which do not appear until adult life; in such cases it may be safely assumed that by the time the characteristic appears the sibship into which an affected person was born will be complete.

Definitions

There will be little difficulty in defining parental age (the age of either or both of the parents at the time of birth of the index case). There are, however, several definitions of birth order, depending largely on the treatment of fetal losses and multiple pregnancies. For example, in calculating the number of previous pregnancies, fetal losses may be ignored, birth order of the patient then being defined as the number of previous livebirths plus one. Alternatively, both liveborn children and fetal deaths may be counted, or, sometimes, only fetal deaths occurring beyond a certain stage of pregnancy may be included with livebirths. Multiple pregnancies may be regarded as contributing either one previous pregnancy or two or more previous children.

The choice of a definition of birth order may be limited by the information available. For example, most national vital statistics record merely the number of liveborn children prior to the birth being reported. The choice may also be influenced by the accu-

racy of the information available. Thus, while the use of the total number of previous pregnancies has some appeal, it has also the disadvantage that differences between groups in the success of efforts to obtain full histories of early fetal loss will be reflected in an apparent difference in birth rank. The choice of a definition may also be influenced by the nature of suspected etiologic factors. For example, if it seems likely that the process of labor or factors associated with the number of children already in the household may have an etiologic influence, then it would seem reasonable to seek for differences between groups with respect to the number of previous liveborn children or of liveborn and surviving children. If, however, the mere fact of pregnancy is likely to have an etiologic influence on the condition, for example, if evidence of hormonal influence is being sought, then the fuller definition of birth order would seem more appropriate. When data are available, it would seem advisable to conduct several comparisons using different definitions.

More important than the choice of a definition of birth order is the requirement that the same definition be applied to all compared groups.

Comparison with the Related Population of Births

Clearly, comparison of the distribution of a series of affected individuals with the related population of births is possible only if that population can be identified — for example, if the affected individuals comprise all the individuals with the particular characteristic identified in a specific geographic or administrative area. In such circumstances the collection of information on the related population may be relatively easy. For instance, several countries, including the United States and Great Britain, publish annual statistics on the distribution of all births by birth rank and maternal age. For the United States, data on age of fathers are also available. Such data allow rates to be computed for any disorder for which information on birth rank or parental age distribution is also available on a nation-wide or state-wide basis. They have been used extensively in studies of fetal death and infant mortality rates, multiple pregnancy, and sex ratio at birth. An example, showing the relationship of fetal death rates to birth order in data for the United States, is given in Figure 34.

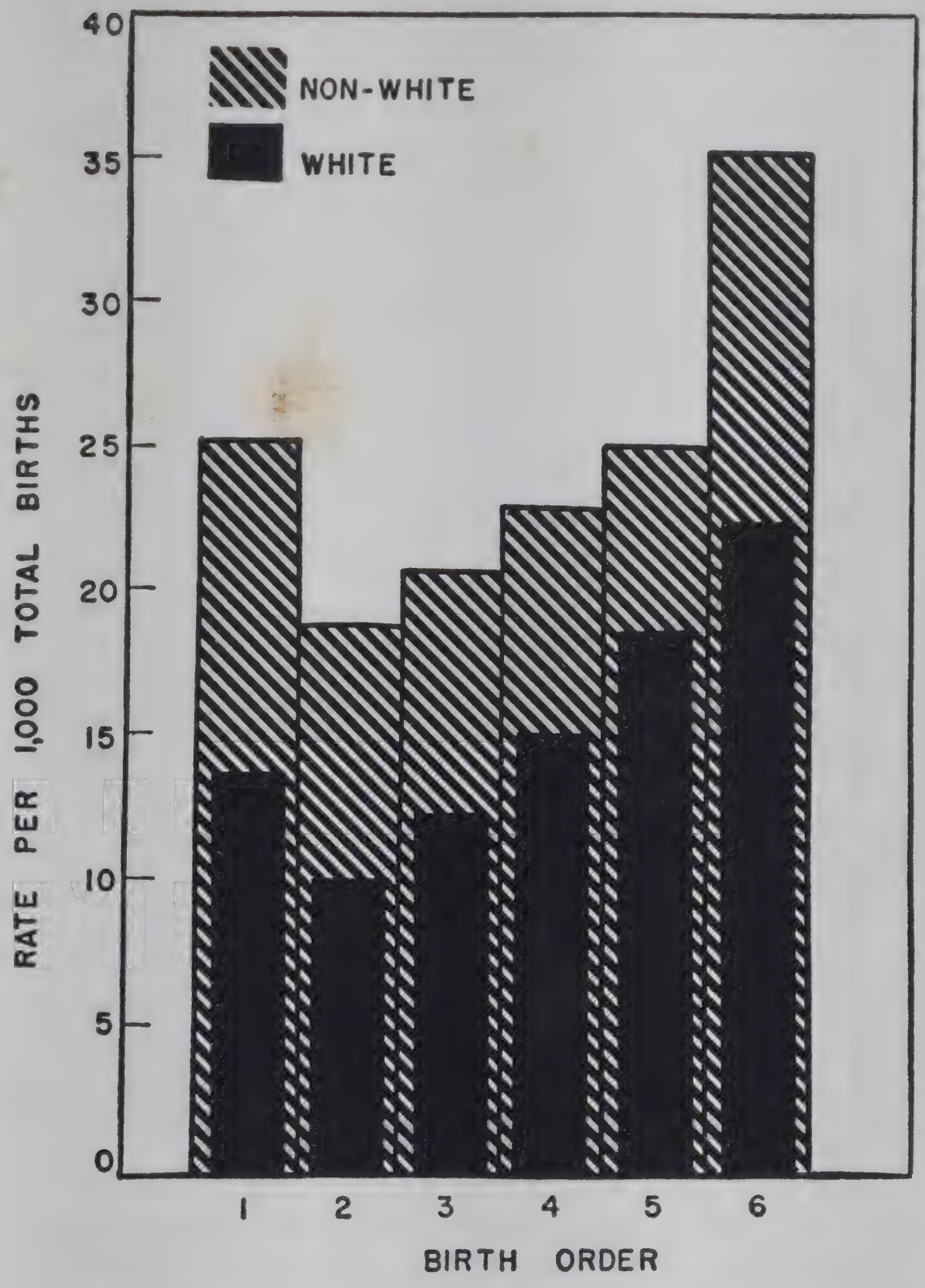


FIGURE 34. Fetal death rates in whites and non-whites related to order of birth, United States, 1957. Birth order is derived from the number of children born to the particular mother, including fetal deaths. Data for this figure and for Figure 36 are derived from Vital Statistics of the United States (National Office of Vital Statistics, 1957). They do not include data for Massachusetts, which does not require the reporting of birth order or birth weight.

Data on distribution by birth order and maternal age of a large population of births are also occasionally available from obstetric records in hospitals where tabulation of this information is undertaken routinely. Material from obstetric records in hospitals has the advantage of detail, of diagnostic accuracy, and of enabling examination of many more disorders and characteristics than are recorded in vital statistics. An example of the use of such material, showing the relationship of the prevalence of mongolism to maternal age, is shown in Figure 35. In areas in which hospitalization for delivery is not routine and where priorities for hospital delivery may be established in part on the basis of age and parity, hospital material may be unsuitable.

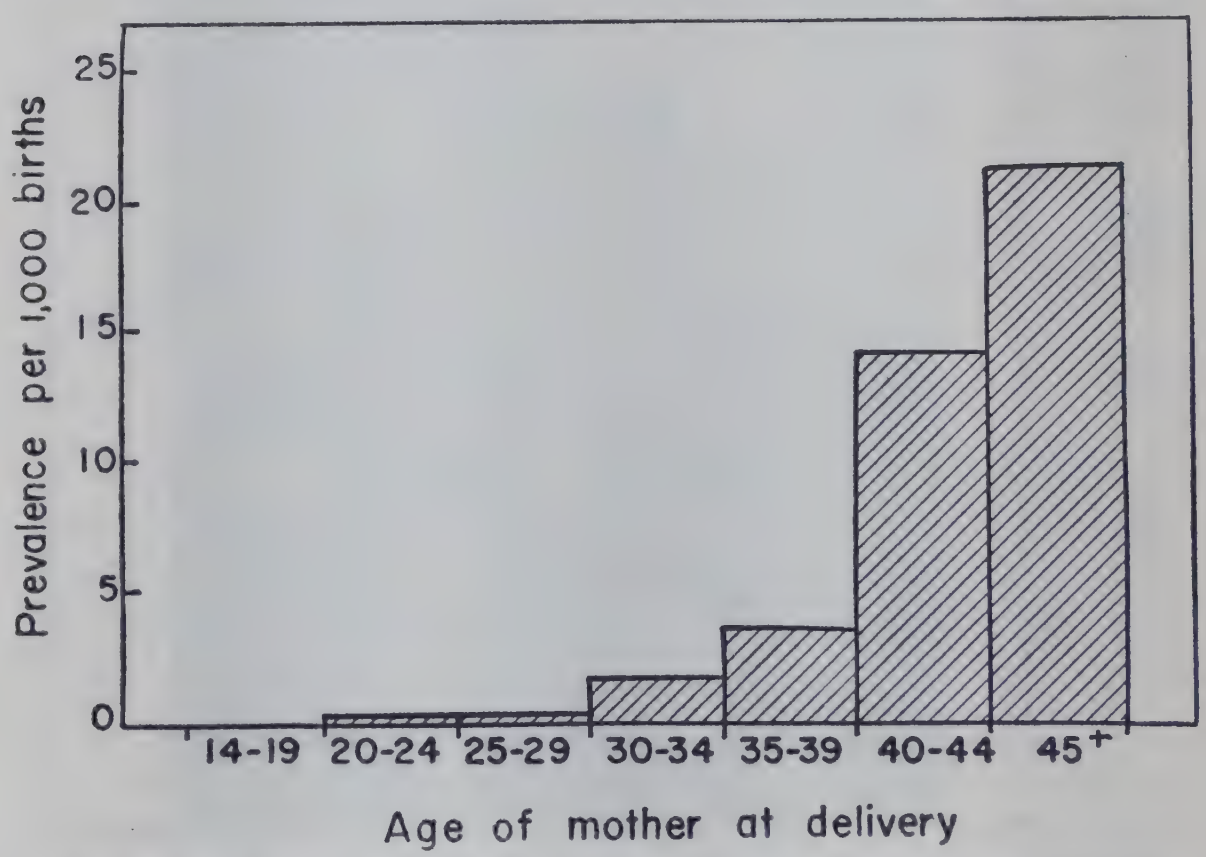


FIGURE 35. Prevalence of mongolism as diagnosed at birth related to maternal age, ten London maternity hospitals, circa 1949. (From Carter and MacCarthy, 1951.)

Use of comparison samples. Comparison with the related population of births may not be possible either because the parental age and birth order distribution of the population is unknown, or because the related population of births cannot be identified. Various forms of comparison samples may then be used. The methods of selection of such samples are similar to those used in

the selection of comparison groups in a wide variety of case history studies and are described at length in Chapter 14. Only a few examples are given here.

When the related population can be identified but parental age and birth order data are not available, the usual procedure is to select a random or systematic sample of the population of births. For example, a number of studies have utilized information derived from a sample of all births occurring in the administrative area of the City of Birmingham, England. This sample, a 1 in 200 sample of all births in the city during a certain period, was originally assembled from registers of births for comparison with a series of infants with malformations of the central nervous system (Record and McKeown, 1949). Subsequently, this sample has been used to investigate the age and parity associations of a wide variety of congenital malformations. In each study the series of affected individuals consisted of all children identified as having the specified malformation who were born in the city during the same years covered by the sample of births. In this case, the required information on birth rank and maternal age was obtained by field inquiry from the mothers in the sample. Frequently however, such information will be available from the sampled records without the necessity of field inquiry. This will almost always be true, for example, if the sample is drawn from hospital records.

If a field inquiry or some further procedure is necessary to obtain information after the sample has been selected, it is important that equivalent effort be devoted to the completion of that information in both the patient series and in the comparison sample. Because of the greater mobility of young adults with small families, that proportion of the sample for whom the subsequent information is not obtained is likely to be over-represented in the younger parental ages and in the low birth orders. Provided this bias affects the two compared groups equally it is not likely to cause serious difficulty. It would, however, be unwise to compare a patient series in which birth order was known for all with a sample of the related population of births in which birth order was unknown for any appreciable proportion.

As described in Chapter 14, some idea of the risks associated with particular birth ranks or maternal ages may be obtained by

assuming that the distribution of the comparison sample represents the distribution of the total population of births. An illustration, using the previously described Birmingham sample, is shown in Table 27. The estimate of the birth rank distribution of all livebirths is derived by applying the percentage distribution of the comparison sample by birth rank to the known total number of livebirths (164,271). These estimates are then used as the denominators of incidence rates.

TABLE 27

*Birth rank distribution of all patients with infantile pyloric stenosis born in Birmingham, England, 1943-1952 and of a 0.5 per cent sample of the related population of births**

Birth rank	Pyloric stenosis series		Comparison sample		Estimated birth rank distribution of all livebirths (164,271 × previous column)	Estimated incidence of pyloric stenosis per 1000 livebirths
	Number	Per cent	Number	Per cent		
1	244	51.1	308	36.1	59,318	4.1
2	139	29.1	271	31.8	52,189	2.7
3	57	11.9	123	14.4	23,688	2.4
4	19	4.0	74	8.7	14,242	1.3
5 +	19	4.0	77	9.0	14,834	1.3
Total	478	100.0	853	100.0	164,271	2.9

* Data from McKeown *et al.* (1951).

If the series of patients to be examined cannot be referred to a specific population of births, for example, in a series of patients with a particular disease assembled from the records of a pediatric hospital, greater difficulty will be encountered in the selection of a suitable comparison group. Some form of paired sample, as described in Chapter 14, will probably be most suitable, although the birth rank distribution of a series of births varies so much in relation to economic status of the mothers, place and time of occurrence of the birth, racial and religious factors, and other variables that it is very difficult to select a comparison group for a particular situation which does not differ from the patient group in a way that might affect its birth rank composition.

For example, Still (1927) published a well-known paper on

"place-in-family" as a factor in disease, noting that the percentage of first-born children among 400 patients with pyloric stenosis admitted to a pediatric hospital was 48.5, compared with a "standard" of 18.5 per cent among patients with lobar and bronchopneumonia, the latter being acute diseases that were considered unlikely to be related to place-in-family. While Still's conclusion — first-born children are more likely to develop pyloric stenosis than are later-born children — was substantially correct (the magnitude of the association in this disease is so great that it is difficult to conceal), the very low percentage of first-born in the "standard" suggests that this was not a suitable comparison group. This is confirmed by the fact that other groups of patients also showed a higher percentage of first-born than the standard but that, for several of these, later work did not indicate the presence of a birth rank association. The nature of the diseases that showed a birth rank distribution similar to the standard (tuberculosis, "digestive diseases") suggests that the standard may have been drawn from a lower economic group than the general run of the hospital's patients.

The Greenwood-Yule Method

Birth order. The basic assumption in the Greenwood-Yule method is that if the risk of appearance of a characteristic is unrelated to order of birth, then among completed sibships of size s containing n affected individuals there will be n/s affected individuals in each birth rank. This supposition is used in the calculation of an expected distribution by birth rank, for comparison with the observed distribution. The method may be illustrated by using the data shown in Table 28. The table is taken from a study by Norton (1952) of 2500 patients with neurosis.

The information required for the calculation of the expected distribution by birth order is the total distribution by sibship size, shown in the right hand column of Table 28. The calculation, in which the data for each sibship size are considered separately, is shown in Table 29. The 226 affected individuals who had no siblings must obviously all be first births. In a random distribution, the 358 affected individuals who came from sibships of size two might be expected to be half (179) first births, and half (179) second births, and the 374 affected individuals in sibships

TABLE 28

*Distribution of 2500 patients with neurosis according to birth order and size of family**

Size of sibship	Birth order of patient																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Total
1	226																	226
2	172	186																358
3	124	128	122															374
4	89	90½†	77½	91														348
5	41	59	72½	73½	89													335
6	36	32	39	37	40	41												225
7	24	26	16	21	24	32	35											178
8	11	11	18	20	17	19	22	21										139
9	11	10	11	7	19	12	16	14	18									118
10	6	4	4	8	13	9	10	6	7	17								84
11	5	1	4	4	3	0	2	3	3	3	5							33
12	0	2	2	1	0	5	1	3	3	3	8	2						30
13	2	4	1	1	2	0	1	2	3	1	0	2	6					25
14	2	0	1	0	0	1	0	1	0	0	2	1	0	1				9
15	1	0	1	0	0	1	0	1	0	0	0	0	0	1	4			9
16	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0		2
17	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	3
18	0	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	3
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
Total	751	553½	369	266½	207	121	87	52	34	24	15	5	6	2	4	1	2	2500

* From Norton, 1952.

† The "halves" are due to the treatment of twin pairs, which were assigned as one-half to each of two consecutive birth orders, the twin pair being counted twice in the calculation of family size.

TABLE 29

*Expected distribution by birth order derived from observed distribution by family size**

Size of sibship	Observed number of affected	Expected numbers of affected individuals occurring at each birth order																	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	226	226.0																		
2	358	179.0	179.0																	
3	374	124.7	124.7	124.7																
4	348	87.0	87.0	87.0	87.0															
5	335	67.0	67.0	67.0	67.0	67.0														
6	225	37.5	37.5	37.5	37.5	37.5	37.5													
7	178	25.4	25.4	25.4	25.4	25.4	25.4	25.4												
8	139	17.4	17.4	17.4	17.4	17.4	17.4	17.4	17.4											
9	118	13.1	13.1	13.1	13.1	13.1	13.1	13.1	13.1	13.1										
10	84	8.4	8.4	8.4	8.4	8.4	8.4	8.4	8.4	8.4	8.4									
11	33	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0								
12	30	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5							
13	25	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9						
14	9	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6					
15	9	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6				
16	2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1			
17	3	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2		
18	3	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	
19	1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Total	2500	794.7	568.7	389.7	265.0	178.0	111.0	73.5	48.1	30.7	17.6	9.2	6.2	3.7	1.8	1.2	0.6	0.5	0.3	0.1

* Data from Table 28.

of size three might be expected to represent the first, second and third birth orders in equal proportions. Thus, the expected distribution by birth order is calculated for each sibship size, and the total expected distribution is derived by summation (bottom row of Table 29).

When the expected distribution by birth order has been obtained, it may be compared with the observed distribution, as in Table 30. Note that, while birth orders may be grouped for the purpose of this comparison, as in this example into "6+," they cannot be so grouped during the derivation of the expected distribution, since the latter calculation requires use of each individual family size. The comparison of expected and observed distributions in this example shows that there was a higher than expected frequency of patients with neurosis in birth orders 5 and 6 and over, and a lower than expected frequency in birth orders 1, 2, and 3.

TABLE 30

*Comparison of total observed and expected distributions by birth order of 2500 patients with neurosis**

Birth order	Observed distribution	Expected distribution	$\frac{Observed}{Expected}$
1	751	794.7	0.95
2	553.5	568.7	0.97
3	369	389.7	0.95
4	266.5	265.0	1.01
5	207	178.0	1.16
6+	353	304.5	1.16
Total	2500	2500.6	—

* Data from Tables 28 and 29.
$\chi^2 = 16.37$, $n = 5$, $p < 0.01$

There are a number of minor points to be considered in these calculations. In the first place, the treatment of twins. Probably the most satisfactory procedure for dealing with twins is to count them as contributing two units to the calculation of family size and, if one of them is affected, to make their contribution to the observed distribution one-half at each of two consecutive birth ranks. Thus, a sibship of size three consisting of one older sibling and a pair of twins, one of whom was affected, would be scored as follows:

	Birth rank			
	1	2	3	Total
Observed	0	½	½	1
Expected	⅓	⅓	⅓	1

The question of inclusion or exclusion of deceased siblings in the calculation of sibship size is of some importance. Neonatal and infant deaths occur more frequently in certain birth orders — particularly the first and the late orders. If the characteristic under consideration is one which is not manifest in early life, there will be a proportionately lower risk of being affected in these birth orders. Consequently, in the calculation of sibship size, only those individuals who lived long enough to have been at risk of exhibiting the characteristic should be included.

Familial disorders. A number of other questions are raised by the application of the Greenwood-Yule method to diseases in which more than one member of a sibship is likely to be affected either because the disease is very common or because it shows a familial predisposition. It is important to note that, in the formula on page 195, n is not the number of sibships in which affected individuals occurred, but the number of affected individuals in those sibships. These two items are the same in the case of conditions that are uncommon and nonfamilial.

When dealing with conditions showing high familial incidence, the error resulting from application of the Greenwood-Yule method to incomplete families will be greater than in nonfamilial conditions, since the probability of the early birth orders escaping and the family being ascertained through an affected individual in the later birth ranks will be decreased.

In a series of sibships, those persons occupying the early birth orders will be on average older than those in the later ranks. When dealing with a familial condition whose incidence increases with age, the possibility that incidence is higher in the early than in the later birth orders simply on this account should be considered. This consideration is not relevant if the probability of more than one affected individual occurring in a family is small.

Incomplete sibships. It has already been stated that application

of the Greenwood-Yule method requires that the sibships examined be complete. Use of the method on a series of incomplete sibships results, as a rule, in an over-representation of affected individuals in the higher birth orders and maternal ages, since data for the later-born members of large families will only be obtained for such families as have an affected individual in the later birth orders. Consequently, there is likely to be an incorrect suggestion of increased risk in the later birth ranks and maternal ages. The shorter the interval between birth and the time of the study, the greater will be the bias from this source.

The extent and types of this error have been considered by McKeown and Record (1956), who list a number of studies in which erroneous conclusions have resulted. These authors also suggest a modification of the method for use when sibships are incomplete and data are not available on the related population of births. The method is suitable for use when cases have been assembled during a limited period, and all siblings born within that same period (either before or after the patients) are known. In this method, the expected distribution by birth order is derived in the same way as in the Greenwood-Yule method, except that siblings born outside the period during which the patients were assembled are excluded. Thus, if the patient series was assembled from affected individuals born during the period 1950 to 1954, a family comprising siblings born in 1949, 1950, 1952 (the patient) and 1955 would be scored as follows:

	Birth rank				
	1	2	3	4	Total
Observed	0	0	1	0	1
Expected	0	½	½	0	1

This method is suitable only in the study of conditions diagnosed at, or shortly after, birth. McKeown and Record explored its application in theoretical and practical situations. The method gives good results on both complete and incomplete sibships in the theoretical models, and in practice does not give the spurious later birth rank effect given by the Greenwood-Yule method on incomplete sibships. However, in a number of actual situations, the method failed to demonstrate a birth order or maternal age

effect known to be present. This failure is attributed to the method being readily disturbed by secular changes in disease incidence or ascertainment occurring over the period of collection of the data. The method does not seem to give sufficiently consistent results to supplant the collection of information on the related population when the latter is possible.

Parental age. Association of a disease or characteristic with age of either or both parents may be assessed by a similar comparison of observed and expected distributions. The expected distribution is derived by apportioning probabilities of the occurrence of the affected individual in accordance with the total number of children in the sibship born at each of the parental ages. Thus a family of four, born when the mother was 22, 26 (affected), 28 and 33 (affected) would be scored as follows:

	Maternal age			
	20-24	25-29	30-34	Total
Observed	0	1	1	2
Expected	½	1	½	2

Consideration of incomplete sibships, twins, and deceased siblings is similar to that outlined in the discussion of birth order.

Separation of Variables

Paternal age, maternal age, and birth order are closely associated. A relationship with any one of them is likely to result in secondary associations with the others. For example, in the study of pyloric stenosis to which reference has already been made (page 194), associations with both birth order and age of mother were found. The data are shown in Table 31. The column totals show that incidence decreased with birth order from 4.3 per thousand in first births to 1.4 per thousand in fourth and later births, and the final column shows that incidence decreased with maternal age from 4.0 in mothers aged 17-21 to 2.6 in mothers aged 37 and over. Either one of these associations could be secondary to the other, since of course first births constitute a larger percentage of all births in younger mothers than in the older age groups. The fact that the decrease is more marked with birth order than with maternal age suggests that birth order may be

the primary variable involved. However, this is no more than a suggestion, since the degree of apparent association will depend in part on the grouping adopted, and the grouping used may be more satisfactory for the one variable than for the other.

TABLE 31

*Incidence of infantile pyloric stenosis per 1000 livebirths, related to birth order and maternal age, Birmingham, England, 1942-1949**

Maternal age (years)	Birth order				Total
	1	2	3	4+	
17-21	4.6	1.8	—	—	4.0
22-26	4.0	2.6	2.9	1.7	3.3
27-31	4.4	2.2	2.2	2.2	2.9
32-36	4.4	3.2	2.5	1.0	2.5
37+	4.3	6.1	3.1	1.2	2.6
Total	4.3	2.8	2.5	1.4	3.03

* From McKeown *et al.*, 1951.

As noted elsewhere, examination of the independent effects of two variables requires examination of association with each of them while the other is held constant. The body of Table 31 shows the trends with maternal age that are evident in each birth order (reading down the columns) and the association with birth order evident in each maternal age group (reading along the rows). When birth order is held constant the inverse association with maternal age evident in the total is no longer seen. On the other hand, the downward trend with birth order is evident in each maternal age group, as well as in the total. This suggests that the association with birth order is independent of maternal age, but that the downward trend with maternal age results from the association with birth order. With regard to the maternal age association, there is, in fact, a suggestion in the data for second births of a maternal age effect directly opposite to that seen in the total for all birth orders.

In addition to the separation of parental age from birth order, there is also the important and even more difficult question of separating the effects of maternal and paternal age. This has been attempted by Yerushalmy (1938) in a study of neonatal mortality, and recently by Novitski and Sandler (1956) in a

study of sex ratio. In both cases, an influence of paternal age independent of maternal age was noted, although in the case of neonatal mortality the association with maternal age was more marked than that with father's age. The Greenwood-Yule method can also be modified to separate the effects of parental age and birth order, as described by Penrose (1934), but the computation is complex and laborious; apart from its initial utilization by Penrose on mongolism, the method has rarely been used.

Effect of Variation in Fertility

Certain forms of variation in parental fertility may produce characteristic birth order or maternal age associations. Particularly, the following possibilities should be considered:

(1) Does the disease under study tend to occur in smaller or larger than average families? In such a situation the Greenwood-Yule method will not suggest an association with birth order unless one is present independently of the association with family size. However, by comparison with the related population of births, a characteristic occurring in small families will appear to be more common in the lower birth orders, and one occurring predominantly in large families will show a positive association with increasing birth order. The patient series may be compared with the related population within given family sizes, to determine whether any association with birth order exists when family size is held constant.

(2) Following the birth of a child affected with the condition under study is there voluntary or involuntary limitation of further reproduction of the parents? Voluntary limitation might, for example, follow the birth of children with certain congenital malformations or follow multiple births, and involuntary limitation might be associated with conditions such as ectopic pregnancy. The effect of such a circumstance on birth order as studied by the Greenwood-Yule method is the same as that of the examination of incomplete families — increased incidence in the later birth orders will necessarily result. Furthermore, since the age of the parents of such families may be such as to ensure that under ordinary circumstances their reproduction would be complete, there may be no way of determining whether a birth order association of this type revealed by the Greenwood-Yule method

results from family limitation or not. The reverse situation in which the birth of an affected individual leads to increased subsequent fertility also produces a spurious birth order effect with the Greenwood-Yule method, this time suggesting increased incidence in the early birth orders. This situation may arise in studies of conditions resulting in stillbirth or early death.

Comparison of affected individuals with the related population of births, or with a sample of the same, eliminates birth order associations arising from changes in fertility subsequent to the birth of the affected individuals, provided the condition under study is not familial. If the condition is strongly familial, family limitation after the birth of an affected individual may result in an apparently increased incidence in the early birth ranks. However, the data collected in a related population group frequently enable comparison of the affected families with the families in the related population with respect to number of children born between the time of birth of the index child and the time of the study. Such a comparison will detect the possibility of this source of error. The data necessary for this examination are not likely to be available in studies of national and other large scale statistics.

INTERPRETATION

If the frequency of a condition is related to birth order or parental age, it may be assumed that some environmental influence is of importance in the etiology of the condition, since the genetic processes of segregation and random assortment are theoretically independent of age and parity, and genetically determined characteristics should therefore be distributed at random to all children of given parents regardless of their birth order or of the parents' ages. Such evidence of environmental influence may alone justify studies of these variables, even if no more specific interpretation is possible, since it may contribute importantly to concepts of the relative roles of genetic and environmental factors in the etiology of the disease in question. Age or birth order associations may be found in conditions in which genetic factors are known to be important, for example, harelip in the mouse (Reed, 1936) and polydactyly in the guinea pig (Holt,

1948). In such cases, however, the existence of environmental influences modifying expression of the genetic potentialities must be presumed. The investigation of the environmental "modifiers" may then have more immediate practical potential than further attempts at elucidation of the genetic make-up.

It should be noted parenthetically that association with birth order or maternal age does not indicate that the resultant environmental influences are necessarily prenatal in timing. Many social and other characteristics of the postnatal and later childhood environments are undoubtedly related to these variables. Naturally any evidence of environmental influence in the etiology of congenital disorders adduced through birth order and maternal age associations must be prenatal by definition.

Types of Mechanisms

Three general types of mechanisms that are explanatory of birth order and maternal age associations may be recognized: (1) nongenic intracellular inheritance, (2) maternal structural and functional changes associated with changes in age and parity, and (3) social determinants of fertility patterns.

Nongenic inheritance. The mechanisms involving nongenic intracellular inheritance are probably the least well understood. In a number of the known examples of nongenic inheritance in the laboratory animal, such inheritance has been shown to be related to age of parents. For example, Lansing (1947, 1954) has shown for the rotifer that length of life of the offspring is related to the age of the parent at the time when the egg is produced. The presence of an "aging factor" which increases in the body of the parent and can be transmitted to the offspring via the egg is postulated. A similar type of relationship has been found by Strong (1954) in studying the latent period of fibrosarcoma development following injection of methylcholanthrene into mice.

No mechanisms of this type have been postulated for man, though average length of life in man has been shown to be shorter for the children of older than of younger parents (Beeton and Pearson, 1901; Dublin and Marks, 1941; Jalavisto, 1950).

Maternal changes. Variation in maternal anatomy, physiology and psychology have evident effects on the intrauterine and postnatal environment of the offspring. This variation may be related

to age and parity. The association of the prevalence of multiovular pregnancy with maternal age must be presumed to have either an anatomic or a physiologic basis. It has been postulated that the development of mongolism may be in part dependent on local structural anomalies of the maternal uterus which are more frequent in older mothers. The evidence to support this explanation with respect to mongolism is meager, but mechanisms of this type are readily conceivable, for example, with respect to syndromes resulting from injury during the birth process.

There are also maternal changes, independent of age, associated with changes in parity, for example, local anatomic changes affecting the uterus. Such changes are likely explanations of the positive association between birth order and birth weight of the offspring. Changes in maternal psychology affecting the postnatal environment of the infant may be related to age or birth order. It seems possible that the high incidence of infantile pyloric stenosis in first births may be explained in this way, since postnatal environmental influences affecting the development of the disease seem to be closely related to maternal care (McKeown, *et al.,* 1955). In this connection, a potentially fruitful field of inquiry would seem to be in more comprehensive examination of the influence of order of birth and parental age in the determination of personality traits.

Social determinants. Social factors productive of birth order and parental age associations are also complex. A distinction must first of all be made between associations which identify the disease as occurring in sibships belonging to a certain (social) type and associations which result from changes occurring within the sibship.

For example, if a cross section of all births occurring in a community is examined, first births will be found to be on the average about a quarter of a pound lighter than second births, which in turn will be lighter than third and later births. This could result from increasing birth weight being directly related to increasing birth order, or from a tendency for women of limited fertility to have smaller babies. In the latter case the early birth ranks would include a high proportion of relatively infertile women, and consequently a high proportion of small babies. The distinction between these two types of relationship can in this

case be made quite readily by examination of the mean weights of first and second births of the same mothers. It has in fact been found that second births are on the average heavier than are the first births of the same mothers, which makes unnecessary the alternative explanation — a possible association between birth weight and infertility. However, examination of successive births to the same mothers is not always easy. So long as a characteristic common to all births (e.g., birth weight) is being considered, numbers of serial observations need not be large, but the difficulty may be insurmountable when a larger number is required, such as in the examination of a specific disease.

A further example of this situation is provided by the fact that patterns of reproduction commonly vary with economic status. Thus, in Western countries, it is common for women in the lower economic groups to reproduce more frequently and at earlier ages than women in the middle and wealthy classes. Consequently in a cross section of births, births in the higher birth orders and lower maternal ages, and particularly births of higher than average birth order for a given maternal age, will be unduly weighted with births to women in the lower economic groups. A disease which is associated with economic status will therefore tend necessarily to show an association with age and birth order. Factors of this type undoubtedly enter into the complex patterns of parental age and birth order associations manifested by fetal and infant death rates, although, as noted above, it is not easy to determine which features of such patterns would be found in an economically homogeneous group of families.

There are also social circumstances which may be peculiar to the sibship. For example, parental social standing, income and child-rearing practices are unlikely to remain static within the sibship as age and parity increase. Further, being born into a sibship already containing children is a determinant of the timing of the common communicable diseases among the later-born siblings.

Much of this discussion of factors explanatory of birth order or parental age associations has been speculative. Although Duncan pointed out the role of maternal age in the frequency of twins as far back as 1865, and Galton (1892) aroused considerable interest in the importance of order of birth toward the end of the last

century, the total amount of statistically adequate work that has been done with these variables has been relatively small and highly concentrated in a few areas such as twins, mongolism, pyloric stenosis and fetal and infant death. The accumulation of information referable to a broader spectrum of outcome will presumably lead to increasing clarification of explanatory mechanisms.

LITTER SIZE

Discussion of prenatal influences would be incomplete without mention of one other variable that may be considered descriptive of the prenatal environment — litter size. That the intrauterine environment is vastly different for twins than for single births perhaps needs no authentication, but is demonstrated by the striking decrease in birth weight associated with increasing litter size (Figure 36). It seems possible that environmental influences capable of producing such striking variation in birth weight may also be capable of producing other characteristics and disorders, particularly those that may arise during the later months of pregnancy when the crowding would appear to be most severe.

The method of determining whether twins are abnormally frequent in a particular disease has usually been to assemble a large series of cases of the defect, to note how many of the affected children were born of a twin pregnancy, and to compare the observed number of twins with an expected number. The expected frequency of a twin confinement among a typical group of white births in the United States is approximately 1 per cent; among a typical group of newborn infants the proportion of twins will therefore be approximately 2 per cent. Adjustment for birth order and maternal age may be necessary, since the frequency of twins among confinements is known to be influenced by these variables. If the proportion of twins among a series of affected individuals differs significantly from that expected, the condition may accordingly be more or less common in twins than in single births. This inference is justifiable when dealing with a series of cases assembled at birth and in which both stillborn and liveborn affected individuals are included. However, since the perinatal mortality of twins is higher than that of single births, some ad-

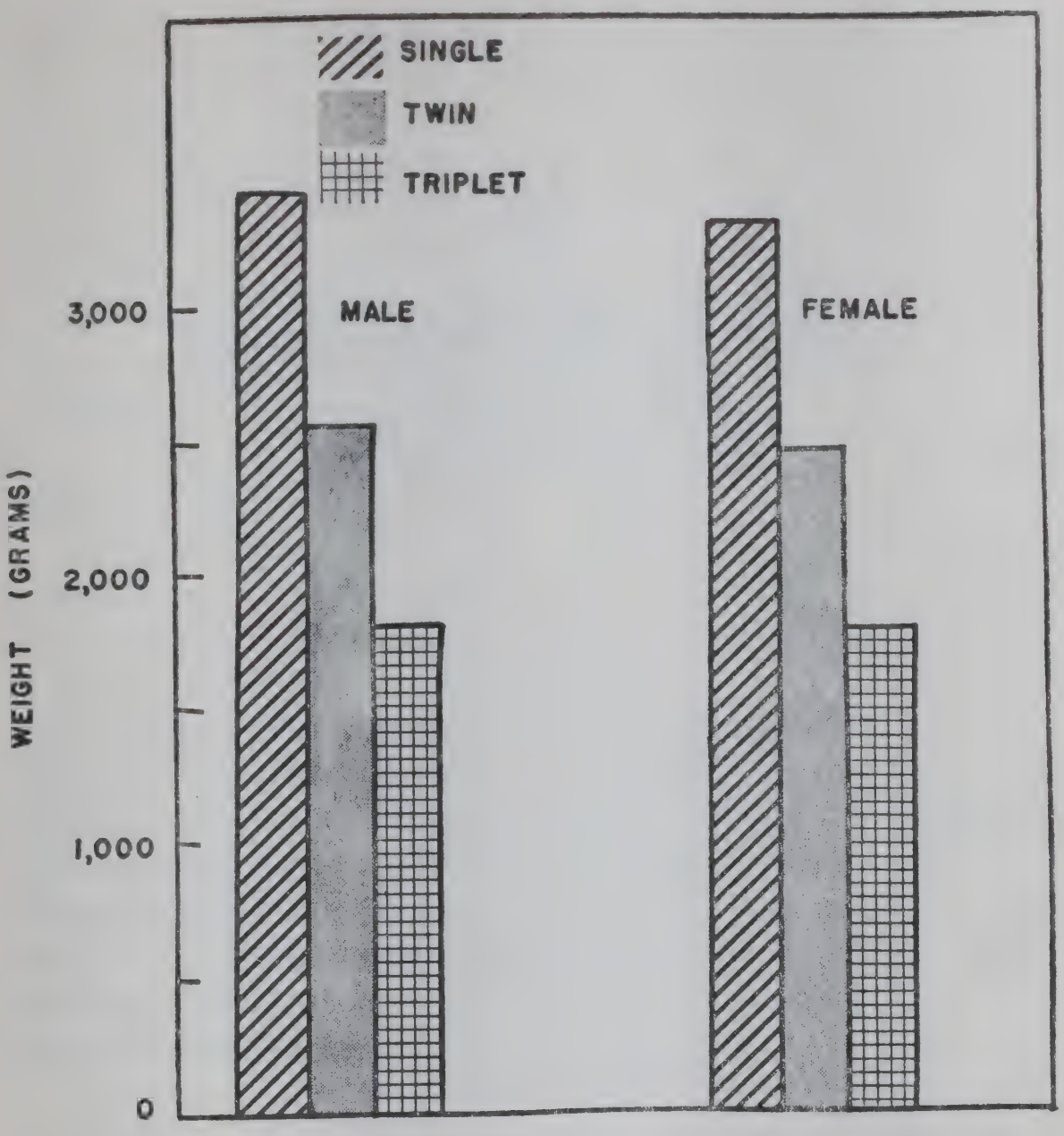


FIGURE 36. Median weight of livebirths from single, twin and triplet pregnancies by sex, United States, white births, 1957.

justment is required if the series is restricted to livebirths or to the neonatal period. Since little is known of the mortality of twins after the neonatal period, adjustment may not be possible if the series is assembled at a time distant from birth. This circumstance has limited the use of this type of investigation almost exclusively to studies of the congenital malformations.

13

Analytic Epidemiology: Cohort Studies

Cohort studies form part of the general area of analytic epidemiology, being studies designed predominantly to test a specific hypothesis or a number of specific hypotheses. As noted in Chapter 4, cohort studies involve the longitudinal study of groups identified in some way other than by exhibiting the characteristic whose etiology is being investigated. These groups are defined through division into those possessing and those not possessing some exposure or characteristic prior to the appearance of the disease or characteristic under investigation.

IDENTIFICATION OF STUDY GROUPS

An important methodologic difference exists between cohort studies in which the study group is the general population, or some readily accessible part of the general population, and those in which the study group is a group identified as having been exposed to some special risk. In the first case, data are required to separate the general population into subgroups according to level of exposure to various factors. The major problem in the second situation is the identification of nonexposed groups whose experience can be compared with that of the special exposure group.

Like most of the classifications of epidemiologic methods that we have attempted, the division of cohort studies into those of

general populations and those of special exposure groups is not invariably clear-cut. For example, the follow-up studies of the populations of Hiroshima and Nagasaki are studies of general populations that might equally well be regarded as special exposure groups. However, since these studies followed a sequence of first identifying a general population and then collecting data to stratify the population according to level of exposure, for purposes of classifying the epidemiologic method involved they may be considered to resemble general population studies more than special exposure group studies.

Studies of General Population Groups

A major difficulty of cohort studies of general populations is the large number of persons who must be followed to obtain adequate numbers of persons developing any particular disease during the course of the study. For example, in order to obtain 50 deaths from coronary heart disease during the course of a year's study, it would be necessary to follow 25,000 persons, on the basis of current death rates for this disease in the United States. Even if the study were restricted to the groups showing the highest rates, say, males aged 35 years or older, it would still be necessary to follow 10,000 persons to obtain 50 deaths. And, it should be recalled, coronary heart disease is one of the commonest disease manifestations in the United States at the present time. For most diseases, therefore, this problem would be much greater. For example, in order to expect 50 cases of leukemia during a year's study, it would be necessary to follow nearly one million people. These numbers can be reduced by longer follow-up, but prolonged follow-up raises its own problems.

Nevertheless, a number of general population studies have been attempted. Some of the different types of groups followed have been:

(1) Whole communities
(2) Occupational subgroups
(3) Medical care subgroups
(4) Insured persons
(5) Self-selected groups
(6) Obstetric populations

(1) Whole communities. To study the occurrence of arteriosclerotic and hypertensive heart disease, a cohort study of the population of Framingham, Massachusetts, was established in 1950 (Dawber *et al.,* 1951). Framingham was, at that time, a town of 28,000 inhabitants in eastern Massachusetts, sufficiently distant from Metropolitan Boston to retain its identity as a separate community and sufficiently large to be expected to yield an adequate number of cases. The study was restricted to the group expected to show the highest incidence — males, aged 30 to 60 years. The study was further restricted to a two-thirds sample of males aged 30 to 60. However, since the sample was intended to be a random one, the study was clearly meant to yield findings representative of males aged 30 to 60 in the whole community. Of the 6510 persons eligible for inclusion in the sample, 4469 were finally included in the study.

The Atomic Bomb Casualty Commission has carried out follow-up studies on the very large numbers of the Hiroshima and Nagasaki populations who survived the atomic bombs of 1945 (Lange *et al.,* 1954). In these studies, 125,041 and 111,135 people in the respective cities were identified; particular interest has centered on the manifestation of leukemia.

(2) Occupational subgroups. The occupations referred to here are not those known or believed to confer special risk of the disease under investigation. They are occupational groups that, with respect to the disease in question, are not known to be unrepresentative of the general population in any relevant respect and which offer special facilities for the collection of data or follow-up. For example, Doll and Hill (1954) studied physicians on the Medical Register of the United Kingdom in 1951. These physicians represented a subgroup of the population who, because of the legal obligation to maintain registration, were relatively easy to identify and trace, who might be expected to receive better than average medical care (and hence diagnosis), and who moreover might sympathize with the objectives of the study. The frequency of lung cancer, the manifestation of interest in the inquiry, was not unrepresentative of that of the general population. Doyle *et al.* (1957) describe a cohort study of male civil service employees in Albany, New York, who because of stability of oc-

cupation and accessibility seemed a suitable group among whom to study the occurrence of heart disease.

Workers in industrial plants also provide populations that are readily accessible and easily followed. However, the number of persons in any one plant is unlikely to be sufficiently large for the study of specific diseases, unless some special risk is present in the plant. The records of unions (particularly those providing medical care programs) may be useful for this purpose, although they have been little used so far.

(3) Medical care subgroups. Persons enrolled in prepaid medical care plans such as union-sponsored medical care programs and the Health Insurance Plan of Greater New York are also more readily followed than the general population. The prepayment component of such plans not only insures that those enrolled will seek the majority of their medical care inside the program, but also provides, through payment of dues, a mechanism for annual follow-up.

(4) Insured persons. The many studies of life insurance statistics might be classified as cohort studies. Persons taking out insurance policies between certain dates constitute the entering cohort, whose subsequent mortality is measured by claims. A special study utilizing such a population group is that of Dorn (1959b) on the effect of smoking on mortality among policyholders of U.S. government life insurance. These policyholders were veterans, primarily from World War I. The cohort consisted of all such policyholders known to be alive in July 1954. This study differs from most insurance studies in that entrance into the cohort was determined not by the time when the insurance was taken out, but by virtue of its members being alive at a specified time, the insurance having been taken out at some time in the past.

(5) Self-selected groups. Although self-selection enters to some extent into all the types of study described above, there are some groups made up entirely of volunteers. For example, Hammond and Horn (1958) assembled a population of 189,854 white men between the ages of 50 and 69 by asking 22,000 American Cancer Society female workers (themselves volunteers) to have a questionnaire filled out by about 10 such men "whom she knew well and would be able to trace."

An attractive feature of self-selected groups such as these is the relative ease of obtaining detailed information, since willingness to supply information is implied by the person's agreement to enter the study. The possibility of bias resulting from self-selection, however, is a major problem. For example, it is possible that the American Cancer Society volunteers, having suspicion that smoking was injurious to health, might have tended to select men who not only were agreeable to entering the study, but who also were heavy smokers and in lower than average states of health. Although individuals who had gross illness were excluded from the cohort, this might not eliminate the difficulty entirely. Such a bias, it may be noted, would be most likely to have an effect early in the follow-up; the longer the interval between the identification of the cohort and the observation of an associated morbid outcome, the less chance there is that the association is due to an initial bias in selection of cohort members.

(6) Obstetric populations. Possible roles of prenatal experiences in the production of defect or disease noted at birth are ideal subjects for cohort studies. In the first place, the period of follow-up is short. Secondly, the delivery and recording of outcome are frequently the responsibility of the same agency (physician or hospital) as that which undertook the measurement of exposure during pregnancy. The two circumstances, indeed, are often recorded in the same medical record.

The Cerebral Palsy Collaborative Study being conducted by the National Institute of Neurological Diseases and Blindness is an extensive study of this type. The cohort consists of approximately 40,000 women registered during pregnancy at one of fifteen hospitals located throughout the United States. The offspring of these pregnancies will be observed for at least the first 6 years of life (Masland, 1958).

Other cohort studies have involved pregnant women registered in individual hospitals and observed at delivery for evidence of fetal defect (McDonald, 1958).

Data on exposure. After the general population group to be studied has been identified, certain data must be collected regarding its individual members. Data will be required:

(1) to classify the group into subgroups according to level

of exposure to the factors or factors whose etiologic influence is being investigated; and

(2) to determine whether the exposure groups so derived are similar with regard to such demographic variables as age, sex, and race which might influence occurrence of the disease under investigation; and

(3) if they are dissimilar in one or more of these respects, to enable allowance for such differences in the analysis.

Such data must be available not only for those who subsequently develop the disease manifestation in question but for the whole cohort. This problem is perhaps simplest in studies using obstetric data, where the records often routinely include data on the occurrence of such events during pregnancy as threatened abortion, hormone therapy, maternal chronic disease, prenatal X-rays, and a variety of demographic data such as race, maternal age, and parity. For the determination of information which would not be expected in routine records, direct access to the pregnant women may be needed. In the Cerebral Palsy Collaborative Study, for example, special interviews with mothers are being conducted together with comprehensive physical examinations and biochemical and other laboratory tests.

When special data are required on less accessible populations the method is more difficult. For example, the cohort studies of lung cancer previously mentioned required detailed information on smoking habits. In the American Cancer Society study, and in studies of insurance data, such data can be obtained from the individuals when they enter the cohort. However, this advantage in ready access to data on level of exposure is more than counterbalanced by the self-selected nature of the group. Doll and Hill (for the British physicians) and Dorn (for the American veterans) obtained the data by mail. Response to a request for information was almost identical in the two studies — 68 per cent of the replies were usable; however, Dorn was able to increase this figure to 83 per cent by a second questionnaire to those who had failed to answer the first.

If data are not obtained for a large proportion of the cohort, the question of bias arises just as in self-selected cohorts. This problem increases as the data of interest become more personal.

For example, in the Framingham population sample certain of the data could only be obtained by physical examination and these examinations were not obtained for 31 per cent of the persons in the sample. The primary reason for failure was lack of cooperation on the part of the person to be examined. Almost certainly the 31 per cent of the sample population for whom the data were not obtained differed from those for whom they were obtained in demographic, psychological and physical characteristics.

The problem of self-selection entering into such studies is usually not as desperate as in those in which the whole cohort is self-selected, since there is often some opportunity to assess the nature of the bias that has entered into the selection and to judge its possible influence on the results obtained.

This assessment consists, first, of comparison of respondents with nonrespondents with respect to such descriptive data (possibly available from other sources) as age, sex and race. This information might be found in available records, for example, of veterans or industrial groups, or might be obtained by inference. For example, if the age distribution of a community's population is known and the age distribution of respondents is subsequently obtained, the age distribution of the nonrespondents can be obtained by subtraction. It is important also to compare respondents and nonrespondents with respect to outcome, if this can be obtained for nonrespondents.

With respect to the yield likely from various sources of information, the Hunterdon County survey is of interest (Trussell *et al.,* 1956). In this survey information concerning medical status was compared according to source. The percentages of persons for whom information was obtained from the various sources were as follows: by interview at home, 91 per cent; by medical examination, 72 per cent; by mail questionnaire, 56 per cent; by multiple screening for chronic disease, 34 per cent. Naturally, such figures will vary according to the information sought, the group in question, and a wide variety of other factors. In one study of a relatively small employment group, 89 per cent of the eligible employees underwent a comprehensive medical examination (Doyle *et al.,* 1957).

Such studies as the one in Framingham, involving periodic in-

terviews and examinations of members of the cohort, allow opportunities for periodic revision of exposure classes according to changes in information concerning its members. In other types of cohort study, such as those of lung cancer, only one assessment of exposure may be made — that at the beginning of the study. Appreciable groups of individuals, however, may change their exposure class between the time of assessment and the end of a long-term study. In long studies periodic reassessment provides a way of overcoming this problem. In any event, it should be noted that changes which occur but which remain unassessed will tend to reduce any actual association that exists between exposure and development of the disease. In other words, the real association is likely to be greater than that found. In fact, a conservative approach, which leans to reduction rather than exaggeration of association, would suggest that it is preferable to utilize the initially determined exposure groups throughout rather than to risk the alteration of an exposure group being influenced by later knowledge of outcome. In studies of exceptionally long duration this generalization may not be applicable. However, in such a situation, separate examination of individuals who have changed exposure groups during the course of the study would still be indicated.

Studies of Special Exposure Groups

Cohort studies of general population groups yield certain information that can be obtained in no other way, and when the factor under examination is one to which a large proportion of the population is exposed (e.g., cigarette smoking), they may be the most economical form of cohort study possible. However, when the factor under examination is one to which only a small proportion of the population is exposed, greatest value for money is obtained by identifying groups who have experienced some unusual exposure and comparing their disease experience with other selected groups.

A number of such studies have been made in recent years of persons receiving large doses of X-rays for medical reasons, for example, ankylosing spondylitis (Court-Brown and Doll, 1957) or thymic enlargement in infancy (Simpson *et al.*, 1955), to determine whether such groups experience greater than average

risks of leukemia and other cancers. The study of Court-Brown and Doll is of particular interest in that the cohort was assembled retrospectively. Assembled in 1955, the cohort consisted of 13,352 patients given X-ray therapy for ankylosing spondylitis in one of eighty-one British therapy centers between 1934 and 1954. The outcome measured was the occurrence of death from leukemia or aplastic anemia at some time between 1935 and 1954. Another example of retrospective definition of a cohort is the study by Case and Lea (1955) of the mortality of a cohort of men defined in 1930 by virtue of receiving pensions for the effects of mustard gas poisoning in World War I (see page 46). This approach adds the major practical advantage of the case history method — immediate availability of data — to the theoretical advantages of the cohort study. It is an efficient and economical method and is being used increasingly.

Other examples of cohort studies of special exposure groups include follow-up studies of infants delivered prematurely and of specially exposed industrial populations such as the investigation by Case *et al.* (1954) of workers in the dyestuffs industry.

Selection of comparison groups. As for cohorts drawn from general populations, it is necessary to assemble certain data on the members of a cohort of specially exposed persons. In the first place, data on age, sex and racial distribution (and possibly other features) are necessary for the interpretation of disease rates. Secondly, while it may be assumed that all members of the cohort are exposed, it may be desirable to attempt a stratification according to degree of exposure. For example, Court-Brown and Doll attempted to determine the dose of X-rays received by the spinal marrow for each of the patients in the cohort with ankylosing spondylitis. The discussion of methods of assembling such data in general population studies applies also to studies of special exposure groups.

Unlike general population studies, special exposure cohorts require the selection of a group of nonexposed persons with which the outcome in the specially exposed group can be compared. The problem is analogous to that of selection of comparison groups in case history studies, although the types of groups selected are generally different. There are two possibilities:

(1) Comparison may be made with rates existing in the general population at the time the cohort is being followed.

(2) Comparison may be made with the experience of another special cohort, similar in demographic characteristics to the exposed group, but not exposed.

In comparisons of the specially exposed cohort with the general population, as done for example by Court-Brown and Doll with respect to risk of death from leukemia or aplastic anemia, rates of the specific outcome in the cohort may be compared with those in the general population. However, since it is nearly always necessary to make the comparison specific for age and sex, most specially exposed cohorts will be too small for the derivation of reliable rates.

It is more usual to compare the observed outcome (say, number of deaths from leukemia) with that which would have been expected in a group of people of the same age and sex distribution as the cohort if the age-and-sex-specific general population rates had prevailed. In the study of ankylosing spondylitis, and in any study in which the study cohort includes all persons with the specific exposure in the population from which the general rates are derived, this is an acceptable procedure; however, when this is not the case, it is necessary to consider whether the manner of selection of the cohort might produce differences in disease rates that are unrelated to exposure. The likelihood of differences arising in this way will vary with the manner in which the cohort is assembled and with the extent to which the frequency of the outcome being measured varies between subgroups of the population.

Comparison with a similar though not exposed cohort might be considered a preferable approach when the exposed cohort does not represent all the exposed persons in a defined population. For example, in industrially exposed persons, persons in other occupations in the same industry may be suitable; Simpson and her colleagues, as well as deriving expected numbers of cases from population rates, compared the frequency of neoplasms in children irradiated for thymus enlargement with the frequency of neoplasms in their siblings; Seltser and Sartwell (1959) compared the mortality of cohorts of radiologists with those of similarly

defined cohorts of pathologists. In theory, an ideal comparison cohort for the studies of patients irradiated for ankylosing spondylitis and thymic enlargement would be patients with ankylosing spondylitis or thymic enlargement who were not irradiated, since such comparison would also rule out the possibility that ankylosing spondylitis or thymic enlargement per se were responsible for any unusual outcome noted. Such groups, however, are often hard to come by.

Studies of cohorts identified at birth offer favorable opportunities for selection of comparison cohorts because of the ready identification of the related population. Thus, cohorts of birth-injured children might be compared with cohorts of infants selected at random or on a paired basis from the same population of births (see page 243).

For purposes of comparison with a specific cohort of exposed persons, it may be that neither general population rates nor another special group may be entirely satisfactory. A study may nevertheless be of value. In the first place, the study may be dealing with persons receiving much larger than average exposures to the agent under investigation, and large increases in incidence of the disease under investigation may be expected. There may be little interest in the demonstration of minor increases. Secondly, cohorts of persons exposed to certain agents are not frequent and all data that can be obtained from them, even if imperfect, may be of value. If firm conclusions cannot be drawn from the results of one study, the sum of several studies of different cohorts exposed to the same agent may be convincing. For example, the results of the study by Simpson *et al.* (1955) of children irradiated for thymic enlargement are shown in Table 32. Comparison with the general population rates can be criticized on the basis of the wide variation in incidence of leukemia and thyroid cancer with regard to other variables (such as geography, ethnic group, and economic group) which were not accounted for in the calculation of the expected number of cases. Comparison with siblings can also be criticized on the grounds discussed in the next chapter (page 238). However, the differences noted are much larger than those that might reasonably be accounted for on the basis of lack of control for these other variables. We may note that the observed cases of cancer, other than leukemia and thyroid cancer,

TABLE 32

*Frequency of leukemia and thyroid cancer among children treated by X-ray in infancy for thymic enlargement**

Disease	Irradiation (1722 children)		No irradiation (1795 siblings)	
	Number observed	Number expected	Number observed	Number expected
Leukemia	7	0.6	0	0.6
Thyroid cancer	6	0.1	0	0.1
Other cancer	4	1.9	5	2.0
All cancer	17	2.6	5	2.7

* From Simpson *et al.* (1955).

exceed the expected number of cases in both the irradiated group and their siblings. This might be interpreted as evidence of more thorough case-finding among the two cohorts than among the general population which furnished the rates for calculation of the expected numbers. However, the extent of this possible bias, as measured in the "other cancer" group, is by no means large enough to explain the relative excess of leukemia and thyroid cancer among the irradiated group. In addition, the consistency of these findings with those of other studies linking increased cancer risk to exposure to ionizing radiation supports the contention that the increased risk experienced by therapeutically irradiated infants, at least with respect to thyroid cancer and leukemia, is quite probably a real one.

FOLLOW-UP

Whether the cohorts to be studied were drawn from a general population or consist of a special exposure group and appropriate comparison groups, there remains the task of determining the outcome among the various groups. In most instances the outcome to be ascertained will be appearance of either morbidity or death from a certain disease manifestation or group of manifestations. It is clear that completeness of such ascertainment should not be unequally affected by membership or nonmembership in a particular exposure category. This objective is most easily attained when the method of ascertainment or follow-up depends

on a mechanism completely separate from that by which exposure was categorized. Perhaps most satisfactory of all is the use of some routine procedure that is undergone by all affected persons regardless of the fact that they are members of a special study group. For example, in Doll and Hill's and in Dorn's studies of cigarette smoking, data on exposure were obtained from the individuals constituting the cohort, but the fact of these individuals dying and the causes of death were obtained from routine records. In the former study, copies of all death certificates on which there was indication that the decedent was a physician were sent to the study group; in the latter, claims for insurance benefits were the primary sources of information. In both, these primary sources were supplemented by inquiry of hospitals or physicians caring for the patient prior to death.

Routine records of death requisite for the termination of a pension, those kept by industry, and those kept by treatment centers were the respective mechanisms of follow-up for the studies of mustard gas pensioners, dyestuffs workers and patients with ankylosing spondylitis referred to earlier. For the patients with ankylosing spondylitis, the routine records of the treatment centers were supplemented by a search of death certificates of the general population for matches between the cohort members and persons who had died of leukemia or aplastic anemia. Such methods are economical since they do not require location and interview of each individual cohort member.

To be efficient the routine source of information utilized must of course be one that will reveal the great majority of the cases of disease or death that actually occur in the cohort. For example, certain of these sources might not provide data on those members of the cohort who, because of emigration, change of occupation, or other reasons, are no longer under observation (even indirect observation). Under these circumstances it may be uncertain how many cohort members actually are at risk of the outcomes under investigation, and the importance of assessing such influences before deciding on this enticingly economical method is obvious.

While satisfactory as a measure of the occurrence or nonoccurrence of such diseases as leukemia and lung cancer, death is often too crude an index of outcome. For example, in the Framingham

study and in the Cerebral Palsy Collaborative Study, indices of heart disease and of infant defect are sought that frequently can only be obtained by specialized clinical examination of living individual cohort members. Such studies call therefore for regular periodic clinical examinations of each member of the cohort. This method yields a great deal more information for the individuals examined than would the use of any routine records. However, it has two disadvantages:

(1) The proportion of individuals examined decreases with the passing years and the membership of the cohort becomes more and more self-selected. The extent and nature of such self-selection can be estimated by comparing cohort members for whom clinical examinations were obtained with those for whom they were not, with respect to their demographic characteristics and crude criteria of outcome not requiring patient contact, such as record of subsequent hospitalization or death.

(2) There is some danger of exposure class and diagnosis of outcome becoming confounded through lack of independency of the observations. This risk might be reduced by such means as not permitting the diagnostician access to past records or by not permitting him to delay diagnostic appraisal. But it is difficult for a physician to remain ignorant of the antecedents of a patient he is examining, or for him to make a diagnosis in the absence of information he considers desirable. The elimination of bias from this source is extremely difficult. "Blind" readings, the use of objective biochemical tests and other measures familiar to the designers of therapeutic trials should be appropriate, but their practicality in this instance has been tested infrequently.

ANALYSIS

Basically, analysis of data from cohort studies involves derivation of rates of a specified outcome among the cohorts under study. These rates can be compared between various exposure groups of the same general population cohort or between exposed and nonexposed cohorts. Alternatively, the number of cases of a disease observed in a specially exposed cohort can be compared with the number expected on the basis of general population data. The requirement that the groups compared be similar with

respect to age and other pertinent variables has been stressed in earlier chapters and is equally applicable here.

Person-Years of Risk

A frequently used denominator for calculating rates of the specified outcome is person-years of risk. This denominator takes simultaneously into consideration the number of persons at risk and the duration of risk of each person. For example, if 10 persons remain in the study for 10 years, there are said to be 100 person-years of risk. The same figure would be derived if 100 persons were at risk for 1 year or if 200 persons were at risk for 6 months. The practical consideration which makes the use of this concept a convenience is that a cohort may not retain the same strength during the whole of the period for which outcome is being measured:

(1) Entrance dates may vary. For example, veterans answering Dorn's second questionnaire on smoking entered the cohort some 3 years after those answering the first. Patients with ankylosing spondylitis entered the British cohort at various times between 1935 and 1954.

(2) During the course of the study certain individuals will drop out from the "at risk" category because of death, loss from the cohort, or other reasons.

In addition to changes in strength, changes in the age distribution of the cohort will necessarily occur as its members are followed in time. If the changes in age occurred equally in the various exposure groups they might, for purposes of comparison between exposure groups, be ignored. However, an assumption that all groups will be equally affected by age changes is rarely a sound one. Separate calculations of risk which take into account differences due to both changing strength and age are nearly always indicated.

Calculation. The data of Doll and Hill (1956), shown in Table 33, are illustrative of the person-years of risk method. No new physicians were admitted to the cohort after the beginning date (November 1, 1951), and the changes in age distribution are therefore due solely to the aging of the cohort and the occurrence of deaths. It should perhaps be made clear that the derivation of the table did not involve repeated censuses of the cohort, but was

based on knowledge of the ages of its members on November 1, 1951, and on the data that were being routinely assembled on its decedents.

TABLE 33

*Number of men at risk according to age at successive anniversaries of entering the study**

Age (years) at specified date	Number of men at risk on						Person-years at risk†
	Nov. 1, 1951	Nov. 1, 1952	Nov. 1, 1953	Nov. 1, 1954	Nov. 1, 1955	Apr. 1, 1956	
Under 35	10,140	9,145	8,232	7,389	6,281	5,779	35,489
35-44	8,886	9,149	9,287	9,414	9,710	9,796	41,211
45-54	7,117	7,257	7,381	7,351	7,215	7,191	32,156
55-64	4,094	4,212	4,375	4,601	5,057	5,243	19,909
65-74	2,694	2,754	2,823	2,873	2,902	2,928	12,462
75-84	1,382	1,433	1,457	1,485	1,483	1,513	6,431
85 and over	181	200	223	256	278	296	1,028
All ages	34,494	34,150	33,778	33,369	32,926	32,746	148,686

* Data from a cohort study of British physicians (Doll and Hill, 1956).
† See text for method of calculation.

The person-years of risk, shown in the last column, were derived as follows: In the age group under 35 years, there were 10,140 men alive on November 1, 1951, and 9145 alive on November 1, 1952. Therefore, if death occurred evenly through the year there were on average 9643 men alive during the first year of the study and these contributed 9643 person-years at risk. Similarly, during the second, third, and fourth years there were an average of 8688, 7811, and 6835 men at risk, respectively. During the last period (5 months) the average number at risk was $6030 \times 5/12 = 2512$ person-years. The person-years at risk during the entire period of observation was, therefore, $9643 + 8688 + 7811 + 6835 + 2512$ or a total of 35,489 person-years.

Calculations made separately for exposure groups enabling comparisons of outcome are shown in Table 34. The summary rate (total) for each exposure group was derived by standardizing rates for each age group to the age distribution of the population of the United Kingdom in 1951.

Difficulties. The major difficulty in the use of person-years is that tests of statistical significance cannot be applied to the resulting rates. The individual person-years constituting the de-

TABLE 34

*Rates of death from lung cancer among British male physicians aged 35 and over, related to smoking habits**

Age group (years)	Nonsmokers			Light smokers†			Moderate smokers†			Heavy smokers†		
	Person-years at risk	Number of deaths	Annual rate per 1000	Person-years at risk	Number of deaths	Annual rate per 1000	Person-years at risk	Number of deaths	Annual rate per 1000	Person-years at risk	Number of deaths	Annual rate per 1000
35-54	11,266	0	0.00	23,102	2	0.09	23,751	4	0.17	15,248	4	0.26
55-64	1,907	0	0.00	6,333	2	0.32	6,514	6	0.92	5,155	16	3.10
65-74	1,078	0	0.00	5,201	7	1.35	3,893	13	3.34	2,290	11	4.80
75 +	856	1	1.17	3,950	11	2.78	1,931	4	2.07	722	3	4.16
Total	15,107	—	0.07	38,586	—	0.47	36,089	—	0.86	23,415	—	1.66

* Data from Doll and Hill, 1956.

† *Smoking categories:* Light: 1 to 14 gm. daily
Moderate: 15 to 24 gm. daily
Heavy: 25 gm. or more daily

nominator do not satisfy the criterion of independence from each other necessary for the routine tests. For example, 10 person-years contributed by 1 individual are likely to demonstrate more similar experiences than the same number of person-years derived from the experience of 10 individuals during 1 year. Furthermore, there is no practical way of determining to what extent complete independence is departed from, and consequently no way of making an allowance for it. In general, therefore, the determination as to whether observed differences based on person-years might be due to chance fluctuations must largely be derived from an inspection which considers the size and pattern of the differences.

Cohort studies do not, of course, necessarily require the use of person-years at risk in the denominator of rates. For example, the use of person-years at risk is not necessary when the period of risk is limited, as for example in studies of pregnancy for fetal defects, when rates can be expressed per person at risk without specification of time (see Chapter 5, page 68).

Persons Lost to Follow-up

In certain types of cohort study, particularly those in which follow-up information is obtained by periodic medical examination or other contact, a number of members of the cohort will be lost to trace at each follow-up examination. Such patients can no longer be considered at risk and some adjustment of the denominator of the rate, whether it is persons or person-years, is required. The difficulty raised by the treatment of persons lost to trace is that the probability of loss may be related to the outcome being measured. Thus, it is possible that cohort members who develop the disease under study may tend to migrate to some other geographic area for treatment, or on the other hand they may tend to move less than those not affected. It is also possible that the probability of loss is related to the exposure group of the person — for example, smokers may be more, or less, "restless" than nonsmokers. Since it is extremely difficult to measure the effect of these considerations, the "only correct method of handling persons lost to follow-up is not to have any" (Dorn, 1950). However, in spite of all efforts that should be, and usually are,

made to reduce their number to a minimum, a hard core nearly always remains.

Ways of assessing the possible influence of losses to follow-up on the findings of epidemiologic studies should be considered in context with the method of obtaining the follow-up information:

(1) When follow-up depends on a regularly scheduled examination in which a change in status between one examination and the next is the outcome being recorded, the most accurate method is to assume that persons lost to trace between two examinations were lost immediately after the first examination. The denominator will then be the number of persons actually examined on each occasion (or the number of person-years experienced between two examinations by persons having the second). The reasoning behind this method is that persons lost between the two examinations cannot figure in the numerator of the rate and thus are inappropriate for inclusion in the denominator.

(2) When follow-up depends on a certain event occurring between two dates and ascertainment of events takes place at the time of their occurrence (as, for example, when cohort members are instructed to report certain symptoms) a number of methods are possible:

(a) If the exact date at which the person leaves the cohort is known, an adjustment can be made for the length of time he was at risk.

(b) If it is known only that a person disappeared at some time between the two dates, it can be assumed that he was at risk for half the period between the two dates.

(c) Two denominators can be calculated — one based on the assumption that all persons were lost immediately after the last date they were known to be present, and one based on the assumption that they were all lost immediately prior to the first date on which they were known to be absent. The correct rate must lie somewhere between the rates based on the two extreme and opposite assumptions.

By use of one or the other of these methods, it is usually possible to make adequate adjustment for the fact that differences in the proportion lost to trace may occur between the various exposure groups. However, the problem of bias resulting from the possible relationship between outcome and probability of follow-

up remains. The only solution to this problem is to make two assumptions, analogous to paragraph (c): first, none of the persons lost to trace developed the specific outcome, and second, all of the persons lost to trace developed the specific outcome. This again yields an estimate of the maximal range due to error of this type. The method may be useful in studies such as those of prognosis in severe diseases in which the proportion lost to trace is small and the frequency of specified outcome (usually death) is relatively high. Unfortunately it has little applicability to epidemiologic studies in which the frequency of the outcome being measured is commonly smaller than the proportion of persons lost to trace, because the range between the two estimates will be too large to have any practical value.

Relative and Attributable Risk

In the analysis there are at least two measures of the strength of association between exposure to a particular factor and risk of a certain outcome. These are:

(1) *Relative risk* — the ratio of the incidence of the disease among those exposed to the incidence among those not exposed. Thus, from the data in Table 34, we could say that the risk of lung cancer was twenty-four times as great for heavy smokers as for nonsmokers (1.66/0.07). This figure is known as the relative risk.

(2) *Attributable risk* — the absolute incidence of the disease in exposed individuals that can be attributed to the exposure. This measure is derived by subtracting the incidence of the disease in question among nonexposed persons from its total incidence among exposed individuals. It is assumed that causes of the particular disease other than exposure had equal effect on the exposed and nonexposed groups. For example, risks attributable to smoking in each smoking category, using the data from Table 34, are shown in Table 35. It is seen that heavy smoking confers a risk of death from lung cancer of 1.59 (1.66 minus 0.07) per 1000 per year; this is 96 per cent of the total risk experienced by heavy smokers.

Similarly an estimate can be derived of the total impact that a specified exposure may have on the community with respect to a particular outcome. For example, the lung cancer death rate

TABLE 35

*Death rates from lung cancer attributable to smoking among British male physicians**

Smoking category	Person-years at risk	Annual death rate per 1000	Attributable death rate†
Nonsmokers	15,107	0.07	0
Light smokers	38,586	0.47	0.40
Moderate smokers	36,089	0.86	0.79
Heavy smokers	23,415	1.66	1.59
Total	113,197	0.81	0.74

* Data from Doll and Hill (1956).

† Death rate in the particular smoking category minus the death rate in nonsmokers.

for nonsmokers (0.07) may be subtracted from the total community lung cancer death rate (0.81); the result obtained might be termed the attributable community risk of lung cancer. The rates for the nonexposed (nonsmokers) and exposed (smokers) must obviously be based on a population similar in demographic composition to the total community. In the illustrative data (Table 35) this was accomplished, at least with respect to age, by the age standardization from which the total rates for each exposure group were derived. The concept of attributable community risk is useful in that it provides an estimate of the maximum reduction of a particular disease rate that might be expected if the specified exposure were removed. On the basis of the data in Table 35, one might expect that 91 per cent of deaths from lung cancer (an annual death rate of 0.74 per 1000) could be avoided if the factor of smoking were eliminated.

INTERPRETATION

One of the difficult problems in interpreting the findings of cohort studies is that measures of relative and attributable risk frequently give different impressions as to the importance of the same exposure, resulting in uncertainty as to which impression is correct. If one is comparing several factors that are etiologically significant in the same disease, the same order of importance of the factors will be suggested whether relative or attributable risk is examined. This is not the case, however, when the significance

of the same exposure is assessed for several different manifestational entities. For example, data from Doll and Hill's study (shown in Table 36) indicate that the relative risk from heavy

TABLE 36

*Relative and attributable risk of death from selected causes associated with heavy smoking among British male physicians aged 35 or more**

Cause of death	Death rate† among nonsmokers	Death rate† among heavy smokers	Relative risk	Attributable death rate†
Lung cancer	0.07	1.66	23.7	1.59
Other cancer	2.04	2.63	1.3	0.59
Other respiratory diseases	0.81	1.41	1.7	0.60
Coronary thrombosis	4.22	5.99	1.4	1.77
Other causes	6.11	7.19	1.2	1.08
All causes	13.25	18.84	1.4	4.59

* Data from Doll and Hill (1956).

† Rates are annual death rates per 1000 related population.

smoking is far greater for lung cancer than for deaths from any other group of causes shown, and at the same time indicate that the attributable death rate is greater for coronary heart disease than for lung cancer. Each of these observations contributes to the interpretation, as follows:

(1) Relative risk gives a clearer idea than does attributable risk of the extent to which the observation explains the etiology of a specified disease. Thus, although the attributable death rate from lung cancer is smaller than that from coronary heart disease, the former comprises a far larger proportion of the total risk of lung cancer (96 per cent) than does the latter of the total risk of coronary heart disease (30 per cent). These percentages are merely different algebraic expressions for relative risk.

(2) Nevertheless, it would appear that heavy smoking causes more people to die of coronary disease than of lung cancer. Consequently, so far as the impact of heavy smoking on a population is concerned, the association with coronary heart disease has at least as great a public health significance as does that with lung cancer.

(3) Both the above arguments depend on the assumption

that the increase in death rate associated with heavy smoking is the result of a causal association. Alternative explanations would be that the increase resulted from some systematic bias inherent in the study method, or that a statistical association existed in fact but resulted from association of both heavy smoking and the specified types of mortality with some unidentified third variable. The size of the relative risk usually is a better index of the probability of causal relationship than the size of the attributable risk. Thus a difference of 10 per thousand (attributable risk) noted between two groups would be less likely to be an error of measurement if it occurred between rates of 1 and 11 than between rates of 110 and 120. It would take less systematic bias to raise a rate from 110 to 120 than to raise it from 1 to 11, just as it is easier to make an error of one inch in measuring a mile than in measuring a foot. Furthermore, the probability of an association between two variables resulting from a chance association of both with a third variable would appear to decrease as relative risk increases, since the relative risk through association with the third variable must be at least as great as that between the first two. The end of the line, however, is not reached until the relative risk attains the value of infinity. Such a situation in practice can obtain only when man, for reasons of utility that are at the same time harmonious with other existing knowledge, constructs by fiat an experiential disease entity (such as tuberculosis) where the relative risk is, perforce, infinity.

Both relative and attributable risk may be seen to play different but important roles in the interpretation of the findings of cohort studies. Although confusion of these roles is to be avoided, there has been no small amount of writing which has pointed to the greater attributable risk of smoking for heart disease than for lung cancer and has then, after proposing alternate explanations of the association with heart disease, implied that similar explanations might account for the larger relative risk in association with lung cancer. Such exercises seem ludicrous and possibly explicable as intentional efforts to cast doubt on an unpopular observation.

14

Analytic Epidemiology: Case History Studies

The place of case history studies in the strategy of epidemiology was discussed in Chapter 4. The term is used to describe studies that are based on the investigation of the past histories of individuals affected with the disease under study. Commonly a specific hypothesis is being tested, for example, that a relationship exists between smoking and lung cancer, or between maternal rubella and congenital malformation. Sometimes, in the absence of leads pointing to specific causal factors, this type of study is used in an exploratory way to investigate the total environmental background of affected persons (to the extent that this can be defined and recalled). As pointed out in Chapter 4, an important component of case history studies is the concurrent investigation of the past histories of a group of individuals who are not affected, to obtain estimates of the frequency of the suspected causes with which their frequency in affected individuals may be compared. Therefore, in the planning and conducting of case history studies, the following must be considered: (1) selection of cases, (2) selection of comparison groups, (3) sources of information, (4) analysis, and (5) interpretation.

SELECTION OF CASES

Two procedures are common in the assembly of a series of patients for a case history study:

(1) The inclusion of all cases of the disease seen between specified dates at some medical care facility or in the practice of a certain physician.

(2) The attempt to include all cases of the disease existing or occurring between specified dates in a limited and defined population, as, for example, in the population of a county, city, or other administrative area.

The first procedure is the more common, since it is relatively easy and inexpensive to carry out. Although the second procedure is considerably more laborious than the first, since it involves special efforts to locate and obtain the necessary data from all affected individuals in the chosen population, it is generally more satisfactory because it avoids the bias arising from the selective factors that lead affected individuals to particular hospitals or physicians, and because it simplifies the selection of a suitable comparison sample.

When a defined population is served by a single medical facility, it may be possible to obtain the advantages of both these methods. For example, in the investigation of congenital diseases or of disorders of pregnancy and parturition, the reference population will be a certain number of pregnancies or delivered births, and the births occurring in a certain hospital or combination of hospitals may constitute the study population. The cases then, although limited to selected medical facilities, will at the same time be drawn from a defined population. Other situations in which a defined population is served by a single medical facility include "closed" populations such as those of schools, service units and work places.

It is preferable that cases included in the series be limited to cases newly diagnosed within the specified study period. Although the inclusion of all cases existing in the period may, in a chronic disease, greatly increase the number of cases available for study, the inclusion of cases at a variety of stages in the disease process may detract from the quality of the histories obtained and compli-

cate the elucidation of time relationships. In patients in advanced stages of the disease it may be difficult to differentiate events causally related to the disease from those consequent to it (as, for example, in a study of the X-ray histories of patients with chronic leukemia). The advantages of restricting the series to new cases is particularly clear when study cases can be related to a defined population, since this will allow estimates of incidence, rather than of prevalence, in relation to the suspected cause.

The establishment of manifestational diagnostic criteria for inclusion and exclusion of potential study cases is an essential part of the design of case history studies. Often it appears wise to include manifestational criteria that embrace possible fringes of the disease "entity" under consideration. This decision is peculiar to each investigation and few generalizations are possible. On the whole, those decisions which lean toward clear and reproducible applications of definition are probably preferable. If time and funds are available, establishment of a number of groups of cases based on variously defined criteria is a happy compromise, providing the possibility of separate examination of each group of criteria. If strong and independent relationships are found, such groupings may assist in judging whether or not certain clinical syndromes should be regarded as "part" of the disease under investigation.

SELECTION OF COMPARISON GROUPS

A comparison group is a group of unaffected individuals believed to reflect the characteristics of the population from which the affected group was drawn. Ideally the comparison group should not differ from the affected group in any respect (other than not being affected) which might be likely to influence the frequency of the variable or variables suspected of being causally connected. This means either that both the patient and comparison groups must be representative of the same population or that if selective factors enter into the choice of the patients, the same factors ought to enter into the selection of the comparison group. Two problems will be considered:

(1) The choice of a large group or sampling frame from which to select the comparison group.

(2) The selection of individual comparison group members from the chosen sampling frame.

The Sampling Frame

The sampling frame for the comparison group usually is determined by the origin of the series of patients. If the patients represent all the affected individuals in a defined population, then the comparison group should also be drawn from that population. On the other hand, if the patients are selected by virtue of attending a certain hospital or physician, a sample of the general population is not likely to be suitable, since factors that lead patients to certain hospitals (such as economic status, race, religion, area of residence, etc.) may also be related to the frequency of factors suspected of causing the disease.

When the origin of the series of patients is not the general population, it appears necessary to find a comparison group that shares as far as possible those selective processes by which the patients came to attention. Three common sources of such groups are discussed in this section, but these by no means exhaust the possible sources of comparison groups. Ingenuity in the selection of new bases for comparison is an important part of epidemiologic study design.

Patients attending the same hospital for other diseases. The rationale behind the choice of this comparison group is that it also is subject to those selective factors that influenced the patients in the study series to come to the particular hospital. For example, if the patients to be used in the study are taken from a hospital located in an area inhabited by a particular ethnic group, a comparison group drawn from other patients attending the same hospital is likely to have the same ethnic distribution as the study group.

It is assumed that the factors that lead a patient to come to a certain hospital are the same for the disease under investigation as for other diseases. This assumption is usually a reasonable one, although not invariably so. For example, if a certain hospital has a wide reputation for the treatment of a specific disease, it might be dangerous to compare its patients with this disease with its general patients, since the patients with the specific disease are likely to be drawn from a wider geographic area and from a

generally higher economic group than those with other diseases. A related point is that, since there may be demographic differences between the patients of different services even within the same hospital, the source of selection of the comparison group within the hospital should be as similar as possible to the source of the study series.

The question will arise whether a comparison group should be drawn from all patients attending the hospital or only from patients with certain groups of diseases that are believed not to be influenced by the factors being studied. For example, if the histories of the smoking habits of a series of patients with lung cancer have been assembled, should these be compared with histories of a sample of all patients in the same hospital or with the histories of a particular diagnostic group, for example, accident cases? The appeal of the second course is that if there are diseases other than lung cancer that are influenced by smoking, the hospital patients will not be representative of the healthy population with respect to smoking, and a difference between lung cancer patients and other patients may reflect the effect of smoking on diseases other than lung cancer. On the other hand, if a diagnostic group can be identified that is known not to be influenced by smoking, the comparison group can be assumed to represent the general population, and a difference between the lung cancer patients and the comparison group can theoretically be presumed to reflect an influence of smoking on lung cancer.

The practical disadvantage of this course is the difficulty of identifying a diagnostic group that is known to be representative of the general population with respect to any particular factor. Consequently, if a difference between the study series and a comparison group drawn from a single diagnostic category is found, it is difficult to decide with confidence which of the compared groups differs from the general population.

On the whole, a comparison group drawn from all patients is probably preferable to one drawn from a single diagnostic category. Since the interpretation of differences between patients in the study group and comparison patients depends to a large extent on the "reasonableness" of different explanations, it is common to use several comparison groups from different sources and to examine the patterns that appear. For example, patients with can-

cer of the lung may be compared with patients with cancer of other sites, patients with diseases other than cancer, and patients without known disease who have come to the hospital for diagnostic procedures. If, in such a study, the lung cancer group differs markedly from *all* the other groups, the evidence is strengthened that it is the lung cancer group whose experience with the suspected factor is unusual. Caution is necessary in the use as a comparison group of patients seen at the hospital in whom no disease is diagnosed. Probably this is the most unrepresentative of any group of patients seen at hospitals.

Relatives of the study group. The advantages of using relatives as a comparison group are the group's close similarities in ethnic and social background to the study group. Two types of relative are used most commonly — spouses and siblings. The use of the spouse may raise difficulties if there is any possibility of a sex difference in the frequency of the manifestation or the suspected cause. It is of course possible to make like-sex comparisons (for example, of male patients with the husbands of female patients), but such comparisons may be impossible or at least inefficient if there is an appreciable excess of patients of one sex in the study group.

When siblings are used as the source of a comparison group, it is necessary to select from among the available siblings on a one patient: one sibling basis. The mistake is occasionally made of comparing a series of patients with all their available siblings. This procedure leads to bias in that the comparison group will be heavily weighted with the members of large families and will show to an excessive degree (by comparison with the study group) any of the multitude of characteristics that are related to family size. For the same reason, patients for whom no comparison sibling is available must be excluded; otherwise the study group, in contrast to the comparison group, will be weighted with one-child families.

The major disadvantage of using relatives for comparison purposes is that their number is limited, and it may therefore not be possible to select a group that is close to the study group in some other important characteristics such as age (see the section on *matching*, page 244). Since many patients may not have any relatives suitable for inclusion in a comparison group, such patients

will have to be excluded, thus removing a proportion of the case material from consideration.

Associates of the study group. Members of the general population may be used on the basis of sharing some common experience with the members of the study group, for example, attendance at the same school, similar place of employment, or residence in the same neighborhood. This type of comparison group combines some of the advantages of the first two types, being comprised for the most part of healthy individuals and at the same time providing a large pool from which individuals with specific characteristics can be selected if necessary. The major disadvantage of the method is the relatively laborious field work that is involved in identification of comparison individuals. In addition, if the information sought is greatly detailed or is of a highly personal nature, an investigator may be reluctant to attempt to obtain such data on a group to which he does not have the entree that is provided by the methods of selection of the first two types of comparison groups. However, there is no doubt that if facilities are available a competently selected group of this type is extremely satisfactory.

Selection of Individual Comparison Group Members

Having selected the source from which the comparison group will be drawn, it is necessary to decide whether the data for the patients in the study will be compared with data on all the individuals who might be available from the chosen source, or whether a sample will be selected. If a sample is decided on, a choice must be made between the various sampling procedures available, the most common of which are random, systematic and paired sampling. It is said that the selection of a source of a comparison group is more important than the method of sampling individuals from that source (Cochran, 1951). This may be true; nevertheless, however satisfactory its source, the usefulness of a comparison group can be seriously affected by unsatisfactory sampling.

Situations in which sampling is not involved. When the general population from which the patients were drawn can be defined and is chosen as the source of a comparison group, it may be pos-

sible to use data on the total population. Clearly this is only practical when data on the frequency of the factors under examination are recorded routinely for the total population.

One of the situations in which total population data commonly form the basis for comparison is in studies of familial occurrence — to answer the question of whether death from the same disease is more common among blood relatives of affected persons than among the general population. Mortality rates observed in the general population are applied to the group of relatives to determine an expected number of deaths which is then compared with the actual number observed. It is of course necessary to consider the number of person-years that the group of relatives has been at risk in each age period, and to apply the age-specific rates in the general population. Depending on the demographic characteristics of the disease being examined, it may also be necessary to apply rates that are specific for other variables, for example, sex and race, and if the disease has shown appreciable secular change, even for the period during which the group was at risk.

In comparing observed and expected numbers of deaths in such a study, it is necessary that the basis of ascertainment be the same. Thus, if the general population rates are based on general mortality statistics, the observed number of deaths must include only those in which the death certificate ascribed the death to this particular disease in such a way that the death would have been so coded in deriving mortality statistics. Additional cases of the disease may of course be uncovered among relatives by virtue of the special investigations undertaken, but they cannot validly be included in a comparison with general population mortality data.

Another common use of population comparisons is the study of maternity hospital data when all births in selected facilities comprise the population at risk. It is becoming increasingly common to process, by punch-card methods, data on characteristics of all births occurring in certain hospitals. Such data may provide valuable material for comparison with data on a series of abnormal pregnancies or births with respect to order of birth, maternal age, blood group, abnormalities of prior pregnancies, and so on. The number of variables that can be examined is again limited by the amount of information that is collected routinely.

In general, medical histories taken routinely in hospitals are

not completed with sufficient consistency to be utilized for study of the types of variables usually examined in epidemiologic case history studies. Therefore, except in the maternity hospital situation, some form of sampling is usually required when hospital patients form the comparison group.

Random and systematic sampling. When collection of data on the frequency of factors under investigation requires extensive record search or inquiry of the patient, it is usual to attempt to assemble such data only for a sample of the comparison group members, rather than for all possible comparison individuals. Random and systematic samples are selected in such a way that the probability of inclusion is the same for each individual in the sampled frame. Such sample groups are therefore similar in composition to the total group from which they were selected. They are used primarily when a sample of a defined population is required for comparison with a series of cases from the same population. Such a sample could also be drawn from certain of the other sources of comparison groups discussed (for example, other patients attending the same institutions), but when samples of such groups are required the method of pairing is usually preferred (see page 243).

A random sample is one in which at each draw each member of the total group has an equal chance of being selected; for example, if each individual is numbered and the numbers selected for study are picked out of a hat or from a table of random numbers. A systematic sample is one in which the group to be sampled is placed in some sort of order and then the selections are made systematically throughout the series, for example, of every second, hundredth, or thousandth individual. Provided the order in which the group is placed prior to systematic sampling is not highly structured with respect to some variable important to the study, the characteristics of a systematic sample are similar to those of a random sample and similar tests of statistical significance may be applied. The distinction between these two methods is therefore not of great importance in most situations, and the choice of one or the other method depends primarily on the practicability of each in a given situation.

The great advantage of random and systematic samples is that, within certain limits of confidence that can be specified statisti-

cally, they allow generalizations about the total group from which they are drawn. As will be pointed out (see page 260), this is particularly valuable when the sampled group is the population from which the patients were drawn. The same confidence cannot be placed in a sample that is selected haphazardly. Bias resulting from haphazard selection is frequent and has been discussed by many writers (White, 1953; Mainland, 1954). For example, haphazard selection from lists of names may be influenced by the length of the name (related to race), its position on the page, or the legibility of the handwriting (possibly related to hospital service and hence to diagnosis). Haphazard selection from files of cards or records may select those which are thicker or more worn and hence cases of greater medical interest.

A special situation that may arise is the necessity to sample from a sampling frame in which certain individuals are listed more than once; for example, it may be necessary to sample from lists of admissions to hospitals, and certain individuals may have been admitted on several occasions. In such a situation a random sample of names from the list will be weighted by persons with multiple admissions. It will be a random sample of admissions, but not of individuals admitted. Moreover, the conversion of such a sample to a sample of individuals admitted requires more than the checking of the selected sample to see that no individual appears more than once in it. It is also necessary to know the distribution of individuals in the sampling frame according to frequency of appearance in the frame, and to adjust the selected sample by making allowance for the probability of the various groups being selected.

Before applying such a correction, it is necessary to be sure that a sample of individuals is required. For example, among a list of births occurring in a hospital over a number of years, certain mothers will appear more than once, according to the number of times they have given birth in that hospital during the period. It is necessary to decide whether a sample of mothers is required (in which case the correction described above must be applied), or whether a sample of births is more appropriate (in which case the correction is not applied). When the series of patients with which the sample is to be compared comprises pregnancies or infants with a certain disease or characteristic, the probability of a

particular woman being included in the patient series is usually related to the number of births she has contributed to the study population. Consequently, if a sample of *women* is selected from the comparison population, it would, by comparison with the patient series, be biased by the inclusion of women of low fertility. In such a situation, therefore, a sample of *births* is the more appropriate comparison group. If, on the other hand, the characteristic under study is a characteristic of the women themselves, it would be more appropriate to select a comparison sample representing *women*.

Paired samples. For the selection of a random sample, the total population to be sampled must be defined and, at least in theory, enumerated. This is not always possible. For example, if the comparison sample is to be drawn from persons living in the same neighborhoods as the patients, how far do these neighborhoods extend? Similarly, if the sample is to be drawn from other patients attending the same hospital as those comprising the study series, it may not be desirable to wait until the total number of such patients is known before beginning to assemble the sample. In these situations, paired sampling is usually undertaken. This involves the selection from the sampling frame of one or more individuals for each patient admitted to the study series. Individuals are selected by virtue of some defined temporal or geographic relationship to a patient, for example, the next patient (or next two patients, etc.) admitted after a study patient, the person living in the nearest residence to that of a study patient, the student next to a patient in the alphabetical class listing, and so on.

Once again it is important that the selection is not haphazard. The method of selection must be rigidly defined and adhered to. For example, in the selection of a neighborhood paired sample it will be necessary to specify whether the interviewer goes to the right or the left of the patient's house, what the procedure is in the case of two-family and apartment houses, and so on. The fact that nobody is at home at the time the interviewer calls is not an adequate reason for rejection of the selected individual, since the probability of all members of a family being out at any one moment is highly correlated with certain demographic characteristics, for example, family size. Whatever the method of selection,

effort must be made to contact selected individuals, the use of "alternate" selectees always raising the question of whether the omission of the difficult-to-locate group makes the sample unrepresentative.

Matching and stratification. In any disease, a series of affected individuals will differ from the population from which it was drawn in many demographic characteristics. It is likely that some of these characteristics are related to the frequency of the suspected causal factors. A difference between the patients and a randomly or systematically selected comparison group in frequency of such factors can therefore be expected, even if there is no causal relationship between the suspected factor and the disease under study. For example, a series of patients with lung cancer will include proportionately more males than will a sample drawn from the general run of hospital patients, and will therefore contain a higher percentage of heavy smokers than the general patient series. In this particular example, the high proportion of males in the lung cancer series is due to the fact that lung cancer affects heavy smokers and consequently males, rather than affecting males primarily and heavy smokers consequently. However, this sequence of events will not be known prior to the study, and it is necessary to examine the possibility that the predominance of males is due to some other cause, independent of smoking, and that the high proportion of heavy smokers is secondary to the excess of males.

There are two general ways of dealing with this type of difficulty:

(1) Statistical techniques used in the analysis of the data that depend essentially on the comparison of like with like — in this example, of males in the patient series with males in the comparison group (see pages 255-259).

(2) Sampling techniques applied to the selection of the comparison group that result in a sample that is more similar to the patient series than would be a random sample. In the present example, a comparison sample could be selected that had the same proportion of males as the patient series; any tendency of males to smoke more than females should then affect both series equally. These techniques are commonly called stratification (when

random or systematic samples are being chosen) and matching (when paired samples are being selected).

The method of stratification requires knowledge of the distribution of both the patient series and the population from which the comparison group is to be selected according to the variables that are going to be equalized. The comparison population to be sampled is divided into subgroups according to the chosen variables, and different proportions are selected from each subgroup. For example, if the patient series contains four times as many males as females, four times as many individuals should be selected from the male subgroups of the comparison population as from the female subgroups.

The use of stratification in case history studies has some considerable disadvantages. In the first place, either because of lack of information or because of mechanical difficulties, it may be impossible or very difficult to subdivide the population to be sampled. Secondly, the necessity to know the final composition of the patient series means that a comparison sample cannot be assembled until after the patient series is complete. This may mean that all the comparison individuals would be interviewed after all the members of the patient series, which would clearly be undesirable. For these reasons, stratification is rarely used in case history studies, although it is common in cohort studies.

When it is decided to select a comparison sample with some of the characteristics of the patient series, it is usually preferable to use the method of paired sampling as previously described (page 243), with the additional requirement that the member of the pair selected from the comparison population must match the corresponding member of the patient series with respect to specified criteria. The selected individual would then be, for example, not the next person admitted but the next person admitted who satisfied the defined criteria.

The decision to select a matched sample or to leave the allowance for differences between patients and comparison group to the analysis depends on a number of factors. In considerable part, the decision depends on what is already known about the characteristic under investigation. Thus, Record and McKeown (1949) in a study of congenital malformations of the central ner-

vous system compared a series of affected individuals with a systematic sample of the related population of births, since little was then known about the association of these disorders with descriptive variables such as maternal age, birth order, season and so on. In future studies of more direct causal influences it might be desirable in the selection of comparison groups to match for some of those variables noted by Record and McKeown and later workers to be associated with these malformations — race, birth order, calendar time, season.

Another factor is the detail of the information sought and the expense of its collection; adequate analysis will require much larger numbers of persons if the comparison sample is unmatched than if it is matched; therefore, if there is any advantage in reducing the number of persons for whom data must be obtained a matched sample is preferable.

Lastly, certain types of comparison groups lend themselves particularly to matching. These include hospital patients, relatives, and neighborhood groups. Since such groups are not representative of any general population, the usual disadvantage of matched groups — that they are not representative of a population and cannot therefore be used as the basis for incidence estimates (see page 260) — is of less significance.

Having decided to select a matched sample, it is necessary to specify the variables that will be matched, and the limits of variation of each variable that will be considered as matching. For example, must age match within 1 year or 20 years? The criteria may be relatively simple. For example, in one case history study of the role of smoking in lung cancer, in which the comparison group was chosen from other patients in the same hospital, the sample was matched by sex and by age within 5 years (Doll and Hill, 1952). On the other hand, quite detailed matching criteria may be set up, as in a current study of the role of X-rays in childhood cancer in which the comparison children, chosen from the neighborhood residents, are required to match for age within 1 year, race, sex, birth order and family size (National Cancer Institute, 1960).

The choice of variables to be matched depends on much the same considerations as enter into the decision to select a matched sample. Since no examination can be made of the possible etio-

logic influence of variables that have been matched, clear knowledge of the role of such variables should be at hand from previous work. Also to be considered is the practicability of the matching process required by the criteria that are set up. The establishment of criteria that are difficult to satisfy leads to loss of patient material (since cases that cannot be matched must be excluded), and perhaps to inaccuracies in the carrying out of the matching process. Three variables that influence the incidence of all diseases and to which a wide variety of historical events are related are age, sex and race. If matching is attempted, it should usually include these variables; beyond these, the matching process becomes more difficult with each additional variable, and each variable added to the criteria requires justification. The criteria for considering two individuals "matched" with respect to any selected variable are determined almost exclusively by practicability. It is desirable to specify limits that are as narrow as possible and at the same time are consistent with one's knowledge of the availability of sufficient individuals to match with the patient series.

One other type of matching should be mentioned. This type occurs when it is necessary to select one individual from a limited number of available persons — for example, when siblings constitute the comparison group. In such a case the selection of one individual from a number of possibilities may be made in terms of an important variable. For example, if a patient has several siblings, the one nearest in age to the patient might be selected, but there would be no specification that the patient and the selected sibling *must* match within so many years of age. While this process tends to reduce dissimilarities between the groups, the patients and comparison relatives will not be completely matched, and it may be necessary to analyze the possible effects of these differences in the production of any associations noted.

SOURCES OF INFORMATION

The common sources of information regarding the factors whose influence is being examined are interviews with patients or with relatives, hospital records, and vital records such as certificates of birth and death.

All these sources of information are open to criticism on one ground or another, and it is fashionable to raise questions concerning the validity of case history studies in general on this account. However, demand for "absolute accuracy" is quite as revealing of a lack of scientific judgment as is intuitive willingness to accept poorly based evidence. Random occurrence of inaccuracy or incompleteness of data should not produce significant differences between study and comparison groups in a case history study. If, therefore, a significant difference is detected, demonstration that the information used was not wholly accurate is not sufficient to dismiss it as meaningless. It must be shown that the inaccuracy was of sufficient magnitude and affected the patient series and comparison group differentially in such a way as could reasonably be expected to produce the observed difference. In data in which no difference has been observed between patient series and comparison group, the effect of inaccuracy of information may be more difficult to assess, since randomly occurring errors and omissions will have the effect of reducing real differences between the two groups and of making the detection of such differences more difficult. In fact, accuracy of information may be more crucial in supporting a negative result than a positive one. In most situations the major issue with respect to accuracy of information is that of obtaining the same level of accuracy in all groups compared.

Naturally, within the limits of the source of information used, efforts should be made to obtain the highest practical degree of accuracy. Each of the sources listed above can yield information of value, but the quality of information will vary from source to source according to the item sought.

Hospital and vital records may be considered together. As noted, neither is a source of the specific information commonly sought in case history studies investigating explicit hypotheses. However, this generalization is subject to modification in the light of local conditions in which a variety of items of information may have been included in routine records for research or other purposes. In many areas maternity hospital records have achieved a degree of uniformity in a variety of data that allow examination of numerous variables and hypotheses.

Data routinely recorded in hospital or vital records may have

advantages over data on the same variables obtained from other sources, particularly if such data have been recorded prior to onset of the disease under investigation and therefore cannot have been affected by knowledge of the disease. For example, in a study of the birth weights of children subsequently identified as being mentally defective, birth certificates or hospital records would be superior to the mothers' memories as sources of information on birth weight not only because of the ordinary deficiencies of human memory, but also because these records have been made prior to the identification of the child as mentally defective.

Many events, for example, maternal infection during pregnancy, if they occur, may be recorded irregularly in hospital records. Some positive histories for such events will therefore be found, but one may be sure that all the positive cases have not been identified. This is an example par excellence of a situation in which a similar degree of inaccuracy must be sought for both patients and comparison group. Greater reliance may be placed on comparisons of frequency of recording of such events if they are recorded without knowledge of the existence of defect — for example, data on maternal infections during pregnancy would be restricted to antenatal records made prior to time of diagnosis of the malformation or other defect under study. Determinations of the presence or absence of a causal factor made after diagnosis of the disease being investigated should be suspect, unless the factor is of a highly objective nature.

There is no more prolific source of data relative to a person's past experience than his or her own memory. The memories of his close relatives may be almost as abundant a source of information and may be available in situations where that of the patient is not, for example, in studies of young children or of persons who have died. The most obvious uncertainty in the use of these sources is that although the human memory may record an event when it occurs, the record is by no means indelible, being liable not only to erasure (inability of recollection) but also to distortion. Moreover the probability of erasure and distortion is related to subsequent events affecting the person as well as to the time elapsed since the event. If a person's memory of trauma to a limb is stimulated by the knowledge that he now has a bone

tumor in that limb, or if a mother's memory of events that occurred during her pregnancy or during her child's early life is stimulated by the fact that her child now has a serious disease, it will be extremely difficult to find suitable comparison individuals in whom a similar degree of stimulation could be expected. Certainly, healthy individuals will not be suitable.

A procedure sometimes employed is to question patients as to the presence or absence of a certain factor and then to check by reference to hospital records or other objective sources those cases in which the factor was stated by the patient to have been present. The sense of security conferred by this procedure is illusory, since in data obtained from patients' memories, errors of omission probably are more frequent and should cause greater concern than errors of commission. A patient will seldom, even in error, report the occurrence of a nonexistent event, but will frequently forget to report an event that actually occurred. This is illustrated in Table 37, which shows the results of a mail ques-

TABLE 37

Hospital record and patients' statements as to the presence or absence of prenatal abdominal X-ray

Hospital record statement	Patients' statement			
	X-rayed	Not X-rayed	Don't know	Total
X-rayed	24	10	3	37
Not X-rayed	2	31	5	38
Total	26	41	8	75

tionnaire to patients whose hospital records had been examined for evidence of their having been X-rayed during a specified pregnancy. Among patients whose hospital records did not indicate that an X-ray had been taken, only 2 patients said that they had had an X-ray. But, among patients who were known to have had X-rays, more than 25 per cent stated that they had not. Checking of the hospital records of the patients who said they had been X-rayed would not in this case have uncovered the major source of error. If patients' statements are to be checked by reference to objective sources of information, both positive and negative statements should be checked whenever possible.

Vagary of memory is only one item among a complex of variables, including bias by the interviewer, the circumstances of the interview, and so on, which may produce differences between patients and comparison group in accuracy of response. The components of this complex cannot always be identified separately. Their effect clearly varies according to the type of information sought, and it is not possible to make generalizations, other than advising the use of common-sense methods in equalizing such bias among the groups being studied. Any technique that can be used to measure the influence of this complex will more than repay the additional labor it may involve.

For example, in a study of 1465 patients with lung cancer, Doll and Hill (1952) noted that the lung cancer patients differed from their matched comparison group (1) by giving a higher percentage of histories of smoking, and in particular heavy smoking, and (2) by reporting a higher frequency of pneumonia and chronic bronchitis in the past. During the course of the same study 335 patients with chest disease were notified to the investigators and interviewed in the belief that they had lung cancer. These latter patients were subsequently found not to have lung cancer and were excluded from the lung cancer series. Nevertheless their histories were compared with those of the true lung cancer patients and with the comparison group (patients with diseases other than cancer). It was found that, with respect to smoking habits, the incorrectly notified group resembled the comparison group and not the lung cancer series, but with respect to history of pneumonia and chronic bronchitis, they resembled the lung cancer series rather than the comparison group. These comparisons suggested (1) that the difference between true lung cancer cases and comparison group in smoking habits was not the result of the patient or the interviewer knowing that the patient had lung cancer, and (2) that there is a tendency for patients with diseases of the chest of all forms to report more pneumonia and chronic bronchitis than do patients with other diseases; this, rather than an association of pneumonia and chronic bronchitis with lung cancer, may be the explanation of the higher frequency of pneumonia and chronic bronchitis observed in the main study.

Other procedures that are appropriate in this connection in-

clude the comparison of data collected by different interviewers, the comparison of histories given by different types of respondents in situations where the source of information is not the same for all patients, and the comparison of data on the same item that might be obtained from more than one source. For example, in a small number of cases Stocks and Campbell (1955) obtained histories of smoking habits of lung cancer patients both from the patients themselves and from their spouses. Appreciable differences were found but these were not sufficiently large to affect the major conclusions of the study. Where practical, provision in a study design for re-interview of a sample of interviewees some time after the first interview is a valuable adjunct.

Nonresponse

Whatever the source of information used, there will usually be some patients for whom the required data cannot be obtained, because the patient could not be contacted, because he or she refused the information, or because the person obtaining the information did not ask the question or did not record the answer. This deficiency of information goes under the general name of nonresponse. It is important for two reasons: (1) in gross amount it reduces the volume of available data, and (2) all groups are not affected equally, the nonrespondents nearly always differing in important characteristics from those for whom the data were obtained. As in the general question of inaccuracy of data, there are two practical considerations; (a) reduction of nonresponse to its lowest practicable level, and (b) determination that the level and type of nonresponse are similar in the study group and the comparison group. The first consideration is perhaps the prior one, since the most satisfactory method of dealing with the second consideration is to eliminate nonresponse in both groups. Even if nonresponse cannot be eliminated entirely, the lower its level, the less its possible effect. However, it is most important that efforts to achieve this minimum be applied equally to both the study and the comparison groups.

For example, Record and McKeown (1949) obtained data by interviews with mothers of children with congenital malformations of the central nervous system and with mothers of a random sample of births. Of the mothers of malformed children 81.2 per

cent, and of the sample mothers 84.7 per cent, were traced and successfully interviewed. For all but 2 per cent of the malformation series data on certain items (e.g., parity) could have been obtained from another source (fetal death certificates), which did not supply data for the sample mothers. In an examination of parity it would not have been legitimate to compare the 98 per cent of the malformation series for whom parity was known with the 84.7 per cent of sample mothers who were interviewed, since the untraced malformation cases included a higher proportion of first births than did the traced cases. This would be expected from the known greater mobility of small families; but, even if such a relationship had not been suspected, the comparison would have been contraindicated. The correct comparison in such a situation is necessarily restricted to the interviewed mothers in both groups.

Some of the variables that can be expected to differ between nonrespondents and respondents in interview situations include the time lapse between identification of the study patients and the attempted interview, age, marital status, size of family, occupation, economic status, and ethnic group. Even nonresponse of a patient due to omissions in hospital records may not occur randomly, being possibly related, for example, to private or ward status, service to which admitted, number of previous admissions, duration of stay, and so on.

ANALYSIS

The basis of the analysis of a case history study is a comparison between study group and comparison group with respect to the frequency of factors whose possible etiologic influence is being tested. This may be a comparison in a fourfold table, such as that in Table 38, in which the concern is merely with the presence or absence of the suspect factor, or it may be in the form of two frequency distributions, as shown in Table 39, where the intensity or duration of exposure is also considered.

In Table 38 it may be seen that a higher proportion of the children with cancer than of the comparison group had a history of prenatal (intrauterine) X-ray exposure suggesting that there may have been an association between X-ray exposure *in utero* and

TABLE 38

*Distribution of 1299 children dying of cancer and of a matched comparison group according to the presence or absence of prenatal X-ray exposure**

Prenatal irradiation	Cancer deaths	Comparison group
Reported	178	93
Not reported	1121	1206
Total	1299	1299

* From Stewart *et al.* (1958).
$\chi^2 = 29.8$, $n = 1$, $p < 0.001$

death from cancer in childhood. Table 39 shows a considerable difference between lung cancer patients and a comparison group in the number of nonsmokers (and hence in the number of smokers); in addition, there is a difference within the group of smokers in the distribution according to number of cigarettes smoked per day. In both these tables, therefore, an appreciable difference exists between the study group and the comparison group. In the further analysis of these differences, at least three distinct questions must be asked:

TABLE 39

*Distribution of 1465 lung cancer patients and a matched comparison group according to average amount of tobacco smoked daily over the 10 years preceding onset of the present illness**

Daily average cigarettes	Males		Females	
	Lung cancer patients	Comparison group	Lung cancer patients	Comparison group
0	7	61	40	59
1-4	55	129	16	25
5-14	489	570	24	18
15-24	475	431	14	6
25-49	293	154	14	0
50+	38	12	0	0
Total	1357	1357	108	108

* From Doll and Hill (1952).
For men: $\chi^2 = 93.77$, $n = 4$, $p < 0.001$
For women: $\chi^2 = 17.41$, $n = 3$, $p < 0.001$

(1) Is the difference statistically significant?

(2) If the difference is statistically significant, can it be explained as a secondary association resulting from other known associations of the disease?

(3) What is the risk estimated to be associated with the presence of the suspected factor?

Statistical Significance

The first of these questions is, of course, the prior one. Unless first shown to be statistically significant, few differences will warrant the consideration outlined below. However, it is possible that even nonsignificant differences, if they satisfy some of the criteria indicated below, may raise sufficient interest at least to stimulate the collection of additional data. It should be remembered that an apparent difference may fail to be significant statistically either because no real difference exists or because the data are numerically insufficient.

It is not our purpose to present material that would be more appropriately found in a textbook of biostatistics, although it must be admitted that in several areas, including the analysis of case history studies, the border between epidemiology and biostatistics is ill-defined. Perhaps the commonest statistical technique involved in case history studies is the chi-squared test for the comparison of distributions. For descriptions of this and other appropriate procedures the reader is referred to such texts of biostatistics as those of Hill (1948), Bancroft (1957) or Croxton (1953). A recent article by Mantel and Haenszel (1959) outlines the specific application of the chi-squared test to case history studies.

Secondary Associations

The second question to be raised is whether the difference between study group and comparison group is secondary to a difference between the two groups in some important characteristic, particularly some major demographic characteristic such as age, sex or race, whose influence on the disease is already recognized. It has already been noted in this chapter that to a certain extent such differences may be reduced in the study design by matching the comparison group to the cases with respect to variables

that are thought likely to complicate consideration of the variables under examination. It will, however, be unusual for all such variables to have been matched, and almost inevitably some analytic procedures will be necessary to examine the effect of variables that were not matched.

There are two problems: (1) the identification of the variables that are likely to produce secondary associations between the disease and the factors under investigation; and (2) examination of the data to determine whether in fact such variables are responsible for the association observed.

The identification of such variables depends in the first place upon knowledge of what is known from previous studies about the natural history of the disease in question and about the distribution of the suspected factor in the general population. Any variable that is known to be strongly associated with either the disease or the suspected factor must be considered as a possible cause of association between the disease and the suspected factor. In addition, the data of the study itself must be examined for evidence of associations that were not previously recognized or that are peculiar to the data of this particular study. This examination should involve two procedures:

(1) Comparison of study group and comparison group with respect to the major demographic variables such as age, sex and race, and such other variables for which data have been collected. Even variables that have been matched in the selection of the comparison group may require comparison if the matching criteria have not been absolute. For example, if the comparison group members were selected by virtue of being within 5 years of age of the corresponding patients, it is conceivable that there may be a consistent bias toward the selection of a comparison group whose members may on average be up to 5 years older than the patients. Where differences are found, the possibility that these reflect conditions that might influence the frequency of the suspected factor must be considered. For example, in the Doll and Hill study shown in Table 39, patients and comparison group members were matched for age and sex. In addition, they were compared with respect to social status, occupation, residence near a gasworks and type of house heating. The fact that the two groups were remarkably similar in all these respects makes the

marked difference in smoking habits more striking, and makes less tenable a hypothesis of the association of lung cancer and smoking being due to a third variable. On the other hand, in the same study differences were noted between patients and controls with respect to frequency of previous respiratory illness, and it was thus necessary to examine the possible effect of this difference on the original hypothesis. In some studies, certain factors are included not for their possible relevance to the etiology of the disease under study, but to enable factual statements to be made about areas of general similarity shared by the study group and the comparison group.

(2) Study of the patterns of distribution of the suspected causal factors separately within subgroups of both the study and the comparison groups. This procedure will produce evidence as to which variables are most likely to influence the frequency of the suspected factor, and should therefore receive most attention in comparisons of patients and comparison group.

When attention has been drawn to a particular variable being possibly productive of secondary associations within the data, it remains to be seen whether the variable does in fact explain the difference between the compared groups that is of primary interest. The statistical procedures involved here may be highly complicated, but the basic examination is a comparison of the study and comparison groups with respect to frequency of the suspected factor within limited ranges of the complicating variable.

For example, Lowe (1956) compared the smoking habits of patients with tuberculosis with those of minor accident patients and general surgical patients, unmatched with respect to age or sex. Since smoking is markedly associated with both these variables, it was necessary to determine whether a difference noted between the smoking habits of the patients with tuberculosis and those of the patients in the comparison group was due to differences in age or sex distribution of the groups. The data are shown in Table 40. For simplicity, only the percentages of smokers of 20 or more cigarettes a day are shown, although, obviously, total distributions according to amount smoked could also be compared in each age and sex category. The table shows that for each sex, in each 10-year-age category examined, the percentage of heavy smokers was higher among the tuberculous than among

the general patients. The difference between the tuberculous patients and the comparison groups in smoking habits cannot, therefore, be explained by differences in their age or sex composition, since the difference in smoking habits remains when age and sex are held constant. Another example of this procedure was given in Chapter 12 (see Table 31, page 202) in the examination of the independent effects of birth order and maternal age on the incidence of infantile pyloric stenosis.

TABLE 40

*Percentage of smokers of 20 or more cigarettes a day among patients with tuberculosis and a comparison group, in age and sex groups**

Age (years)	Number of patients				Percentage smoking 20 or more cigarettes a day			
	Males		Females		Males		Females	
	Tuberculous	Other	Tuberculous	Other	Tuberculous	Other	Tuberculous	Other
30-39	211	126	130	113	48.3	46.0	9.2	2.7
40-49	164	141	36	94	51.8	44.6	19.4	3.2
50-59	104	103	19	42	51.0	41.8	10.5	0.0
60+	41	49			51.2	36.8		
30+	520	419	185	249	50.1	43.4	11.4	2.4
30+†	—	—	—	—	50.3	43.7	12.4	2.5

* Data from Lowe (1956).

† Standardized to the age distribution of the combined cases and comparison group.

In conducting examinations such as these, the method of grouping is important. The subgroups set up must contain a sufficient number of cases for separate examination of each subgroup, but they must not be so broad that the desired similarity between compared series is not obtained. For example, if a disease is markedly related to age, the use of 20-year-age groups may not eliminate the effect of age, since within 20-year-age groups there may be sufficient difference between two series to produce appreciable differences in disease incidence. It may be difficult to decide whether the adopted grouping has been sufficiently fine to effectively equalize the compared series. If the apparent difference between compared groups is eliminated by holding another variable constant, it may be assumed that the grouping was satisfactory. How-

ever, if the apparent influence of the variable is reduced but not eliminated, the effect of finer grouping should be ascertained. If finer grouping is impractical because of inadequate numbers, it may be possible to estimate by indirect methods the maximum possible influence of variation within the chosen groups by making assumptions about the maximum extent of variation possible within the adopted grouping and the maximum change in incidence that might be associated with it. Some general idea about the possible effect of finer grouping may also be obtained by noting the change in the crude data that results from examination of relatively broad groups, as discussed in the next paragraph.

In addition to visual inspection of the differences between the study and comparison groups in each subdivision of the data, it is usually desirable to test the statistical significance of the differences and to obtain some sort of summary statement that will express the sum of the differences noted in each subdivision — in other words, the total difference between the study and comparison groups that exists independently of any difference attributable to the complicating variables. The significance test that will be most often appropriate is the cumulative chi-squared test, as described in relation to this problem by Mantel and Haenszel (1959). A simple statistical technique that may be used to obtain a summary statement of the difference between groups after the effect of a complicating procedure variable has been eliminated is one or other of the standardization techniques that are commonly used in adjustment for age differences between compared tabulations. Standardization of the data in Table 40 to the age distribution of the combined cases and comparison group results in the percentages of smokers of 20 or more cigarettes a day shown in the bottom row of the table. These figures are little different from the crude totals before standardization, which suggests that an even finer subdivision of age groups than the 10 year periods used would not appreciably affect the results.

Estimated Risk

The third question in the consideration of a difference noted between the study and comparison groups concerns the size of the difference — how great a risk is apparently conferred by exposure to the factor for which the difference is noted. In the previous

chapter, it was noted that such a risk may be expressed in two ways:

(1) Relative risk — the risk (rate) in exposed persons divided by the risk in the nonexposed group.

(2) Attributable risk — the risk in exposed persons minus the risk in the nonexposed group.

In estimating these risks in case history studies it is necessary to distinguish clearly between studies in which the cases can be related to a defined population and those in which the cases have been assembled without reference to a defined population.

Studies of cases referable to a defined population. When the cases represent all cases of the disorder in a population or are a sample of such cases selected in a known way and the comparison group is representative of the same population, it will be possible to estimate separate rates of occurrence of the disease in exposed and nonexposed persons and to derive relative and attributable risks from these estimates.

For example, the cases of childhood cancer shown in Table 38 (page 254) were 77 per cent (all those for whom data could be obtained) of the 1694 children under 10 years of age who died of cancer in England and Wales in the years 1953 to 1955. The comparison children were selected from birth certificates in the same locality in which the deaths of the cancer patients occurred and were matched for age and sex. The assumption may be made that, with respect to frequency of prenatal X-ray irradiation, the comparison group was representative of the total number of births that resulted in the population of children aged less than 10 years in the period 1953 to 1955. Since 7.2 per cent of the comparison sample population received prenatal X-rays, it is assumed that a similar proportion of all children alive in the population at risk received prenatal X-rays. The average population of England and Wales aged less than 10 years in 1954 was approximately 6,999,000. The population at risk may therefore be supposed to consist of approximately $6{,}999{,}000 \times \frac{7.2}{100} = 503{,}900$ children who were X-rayed prenatally and $6{,}999{,}000 \times \frac{92.8}{100} = 6{,}495{,}100$ who were not. By applying to these populations the numbers of cancer cases found to have been X-rayed (178) and those not X-rayed (1121),

one can derive death rates in the X-rayed and the non-X-rayed population during the 3 year period. Division by three yields average annual death rates of $\frac{178}{503{,}900 \times 3} \times 10^6 = 118$ per million in the X-rayed, and $\frac{1121}{6{,}495{,}100 \times 3} \times 10^6 = 58$ per million in the non-X-rayed children. If it is assumed that the 395 untraced cancer cases were distributed by X-ray status in the same way as the traced cases, these rates become 154 and 75 per million, respectively.

Studies of cases not referable to a defined population. The data in Table 39 (see page 254) cannot be related to a population. The lung cancer patients were patients attending selected hospitals and no attempt was made to limit the group according to place of residence or to determine what proportion these represented of the total lung cancer patients in the areas served by the hospitals. Similarly, the comparison group was selected from patients with other diseases in the same hospitals, and although it was therefore a sample of the same population from which the lung cancer patients were drawn, it is not known what proportion of that population they represented.

Since neither the population at risk nor the numerator of the rate is known, it is impossible to derive separate estimates of incidence in the exposed and the nonexposed populations; it is therefore impossible to derive the risk attributable to exposure. However, it is possible to derive a relative risk for the following reasons. It was noted in Chapter 4 that a theoretical population can be divided among the cells of a fourfold table as follows, where a, b, c and d represent the frequencies of individuals in the various cells.

Suspected cause	Disease		Total
	Present	Absent	
Present	a	b	$a+b$
Absent	c	d	$c+d$
Total	$a+c$	$b+d$	1

In a case history study, only derivatives of such frequencies are available, and the numbers of individuals in each cell may be denoted as a', b', c', and d'.

Historical episode	Cases	Comparison group
Present	a'	b'
Absent	c'	d'
Total	$a' + c'$	$b' + d'$

The right-hand column has been omitted since, as noted, it cannot be derived in a case history study unless it is known what proportion of $(a + c)$ and $(b + d)$ are represented by $(a' + c')$ and $(b' + d')$ — in other words in what proportion to weight the cases and comparison group columns to derive an estimate of total population.

From the first diagram showing the theoretical distribution of the population, one can see that the risk in exposed persons relative to the risk in those not exposed would be:

$$\frac{\text{incidence of the disease in exposed persons}}{\text{incidence of the disease in nonexposed persons}} = \frac{a}{a+b} \div \frac{c}{c+d} = \frac{a(c+d)}{c(a+b)}.$$

When the number of persons affected by the disease under investigation is small compared to the number unaffected (the usual situation in disease studies), d is approximately equal to $c + d$, and b is approximately equal to $a + b$. The formula for relative risk then reduces to $\frac{ad}{cb}$.

When the cases studied represent only a sample of all cases in the population, and the comparison group is a sample of the population at risk, then $a' = as_1$, $c' = cs_1$, $b' = bs_2$ and $d' = ds_2$, where s_1 is the sampling fraction used in selecting the cases and s_2 is the sampling fraction used in selecting the comparison sample from the population. The formula becomes:

$$\text{Relative risk} = \frac{a's_1d's_2}{c's_1b's_2} = \frac{a'd'}{c'b'}$$

It is clear therefore that for the derivation of relative risk it is not necessary to know the values s_1 and s_2.

For example, from the data in Table 38 (see page 254) one can calculate that the risk of death from cancer for children reporting prenatal irradiation was $\frac{178 \times 1206}{93 \times 1121} = 2.0$ times as great as for

those not reporting such irradiation. This is consistent with the estimates calculated earlier in the chapter of death rates of 154 and 75 per million in exposed and nonexposed children, respectively. From the data in Table 39 in which the values s_1 and s_2 are not known, the relative risk for smokers (any amount) compared to nonsmokers would be $\frac{1350 \times 61}{7 \times 1296} = 9.4$.

The use of the above formula involves two assumptions: (1) that the disease under study is infrequent; and (2) that neither the patient series nor the comparison group is selected in favor of either exposed or nonexposed individuals — that is, that the fractions s_1 and s_2 apply equally to exposed and nonexposed individuals in the cases and the comparison group, respectively. If the first of these assumptions is open to question, the formula still has use since the estimate derived is a function of relative risk.

If the design of the study is inconsistent with the second assumption, however, a derived relative risk is necessarily misleading. For example, the relative risk of lung cancer in relation to smoking has been found to be generally smaller in studies in which hospital patients formed the comparison sample than in cohort studies in which the comparison group derived from the general population. Menaker (1956) has pointed out that the reported proportion of nonsmokers among hospital patients is smaller than among the population at large, presumably because a number of diseases other than lung cancer are positively associated with smoking. Thus, in case history studies in which hospital patients are used as a comparison group, the proportion sampled from the general population may be higher for those exposed (smokers) than for those not exposed (nonsmokers). In this case the error is in favor of understatement. The relationship to smoking is so much stronger for lung cancer than for other diseases that the bias did not obscure the relationship. However, under other circumstances such a bias could readily operate in the other direction.

Adjusted relative risk. We have considered a relative risk derived from a simple fourfold table containing a summary of the study data. This must be considered a crude relative risk, since no allowance has been made for the fact that some or all of the apparent association noted may result from indirect associations with

variables such as age, whose influence has not been considered in the total data. As noted previously, the possible effect of such variables must be examined by classification of the data to allow comparison of cases and comparison group within defined limits of the complicating variable. Within each subclassification a relative risk can then be calculated by the formula $\frac{a'd'}{c'b'}$.

For example, in Lowe's data for males shown in Table 40 (see page 258), the relative risks for smokers of 20 or more cigarettes a day are 1.10, 1.33, 1.45 and 1.80, respectively, in the four age groups shown. The problem then is to derive an over-all relative risk which is a summary of the relative risks estimated within each of the subclassifications. One procedure is to apply the formula to the total distribution after standardization. For example, using the standardized data for males in the table, the relative risk is $\frac{50.3 \times 56.3}{43.7 \times 49.7} = 1.30$.* In this particular case, the effect of standardization is small, the relative risk being practically identical with that which might be calculated from the total crude data for males — 1.31. The estimated summary relative risk may vary somewhat according to the population used for standardization. Other methods of deriving summary relative risks have been described by Mantel and Haenszel (1959).

Multiple grouping of study variables. We have so far considered only situations in which the classification of the variables under study is dichotomous, such as in Table 38 and Table 40. Clearly, the setting up of an artificial dichotomy at a certain level of exposure such as 20 cigarettes a day, as in Table 40, underestimates the relative risk that is associated with the factor considered. A more realistic estimate of the importance of the factor is obtained by comparing those experiencing minimal exposure with those maximally exposed, for example, in Table 39, by comparing nonsmokers with those smoking 50 or more cigarettes a day.

The formula $\frac{a'd'}{b'c'}$ can be applied separately to any exposure level — it is not necessary that c' and d' be the total number of cases minus a' and b'. For example, for the data for males in Table 39,

* Note that the formula can be applied to percentages as well as to actual numbers.

the relative risk of lung cancer in smokers of 1 to 4 cigarettes compared to nonsmokers is $\frac{55 \times 61}{7 \times 129} = 3.7$. Likewise, the relative risk to smokers of 5 to 14 cigarettes compared to smokers of 1 to 4 is $\frac{489 \times 129}{55 \times 570} = 2.0$. A series of relative risks can therefore be derived for adjacent categories of smoking. These are shown in the column labeled (*a*) of Table 41. In this particular study it adds greatly to the strength of the hypothesis that the relative risk is greater than 1 in each comparison of a category of smoking with the lower adjacent category.

TABLE 41

Relative risks as derived from the data for males in Table 39

Daily average (cigarettes)	Lung cancer patients	Comparison group	Relative risk	
			(a)*	(b)†
0	7	61	—	1.0
1-4	55	129	3.7	3.7
5-14	489	570	2.0	7.4
15-24	475	431	1.3	9.6
25-49	293	154	1.7	16.3
50+	38	12	1.7	27.7

* Relative to persons in the adjacent lower category of smoking.
† Relative to nonsmokers.

Since smokers of 1 to 4 cigarettes have 3.7 times the risk of nonsmokers, and smokers of 5 to 14 cigarettes have 2.0 times the risk of smokers of 1 to 4, it follows that smokers in the 5 to 14 category have $2.0 \times 3.7 = 7.4$ times the risk of nonsmokers. Therefore, by accumulative multiplication, a series of risks in each exposure group can be calculated relative to a risk of 1.0 in the lowest group. This series is shown in the table in the column labeled (*b*). The risk of 27.7 for the highest exposure group (50 or more cigarettes) relative to the lowest (nonsmokers) could also have been derived by direct calculation for the two groups $(38 \times 61)/(7 \times 12)$.

In further illustration of the effect of using a dichotomy, as, for example, at 20 or more cigarettes a day in the data for respiratory tuberculosis in Table 40, we may note that, compared with a relative risk of 1.3 derived by comparing smokers of fewer than and more than 20 cigarettes a day, in the same study the relative risk

in the highest exposure group (40 or more cigarettes a day) compared to the lowest (nonsmokers) was 5.3.

INTERPRETATION

At the termination of a particular case history study, it will be possible to determine whether or not in the data of the study in question a significant association exists between the variables pertinent to the hypothesis being tested and the disease under investigation. The attainment of this objective, however, serves a useful purpose only to the extent that the findings of the particular study are indicative of the existence of an association between the variables and the disease capable of generalization to all situations similar to those of the study. Such generalization appears most permissible when the study has been designed and conducted in such a manner that it seems reasonable to assume the following:

(1) The cases studied are indeed representative of cases of the disease in question.

(2) The comparison group reflects with accuracy the relevant characteristics of the general population.

(3) The sources of information indicate the past experience of the cases and comparison group members with a high degree of precision.

(4) The observed differences between cases and comparison group have been analyzed with sufficient rigor to indicate their probable causal as well as statistical significance.

It is not, of course, necessary that these assumptions be accepted or rejected *in toto*. As pointed out previously, data containing errors or biases need be rejected only if such errors or biases are of sufficient magnitude and operate in a direction which could explain the observed patterns. Similarly, no series of cases can be said to be fully representative of all cases of the same disease, but in many situations a specific series will be sufficiently representative to allow conclusions in specific areas. Likewise, no comparison group will be truly representative of the general population, although in a well-designed study it will not be unrepresentative in any relevant respect.

In addition to assessment of the reliability of the data from a

particular study, a number of more general questions enter into the interpretation of the findings. These include the way in which the observations fit into the pattern of existing knowledge of the disease, the consistency of the observations between subgroups of the data (e.g., is the observed relationship observed in both males and females?), and the extent to which the observations are confirmed in other studies. A new observation need not necessarily be explainable through mechanisms consistent with existing knowledge, but most new observations that have proved useful have not seemed ridiculous in the light of existing knowledge. Also, existing knowledge helps to put a newly observed relationship in perspective. For example, a relationship responsible for doubling the risk of a disease will have limited significance if much stronger relationships with other variables are already known. On the other hand, the minor difference in the proportion of males among first births (51.7 per cent) and among sixth and later births (51.3 per cent) is of considerable interest, because stronger relationships to the sex ratio of human births are not known.

Perhaps the most convincing evidence for the ability to make a generalization from an observation comes from replication of results by other studies. When varied methodologies result in similar conclusions, increasing confidence can be placed in generalization from the observations.

An additional purpose in case history studies will usually be to determine the likelihood that any statistical associations noted are causal in nature. (For a discussion relative to these inferences see Chapter 2.) However, it may be noted here, that, despite many statements to the contrary, the difficulties of inferring cause from observed association are not peculiar to case history studies; they are equally evident in cohort studies. The specific difficulties, as indicated in Chapter 2, are characteristic of the general field of observation (as opposed to experiment) and are to be found in all observational studies, regardless of method.

15

Experimental Epidemiology

Up to this point we have been dealing with observational methods exclusively. As noted in Chapter 2, the most rigorous test for causal association is that of experiment — the artificial introduction of an attribute or treatment in an attempt to alter the course of nature. Such experiments in the field of the natural history of human disease range from field trials of vaccines and other agents in the prevention of disease to clinical trials of therapeutic procedures that hopefully have a favorable influence on the course of disease after it has arisen. Usually, the epidemiologist is more concerned with experiments dealing with etiology rather than therapy, and with trials involving communities or large populations rather than those restricted to ill persons in hospital beds. Our remarks are therefore concentrated in the area of trials of preventive measures.

UNCONTROLLED EXPERIMENT

It was noted in Chapter 4 that the application of any public health measure may be considered an experimental test of the validity of knowledge of disease etiology. Figure 37 demonstrates the trend of notifications of diphtheria and whooping cough in Massachusetts during this century. Diphtheria immunization was

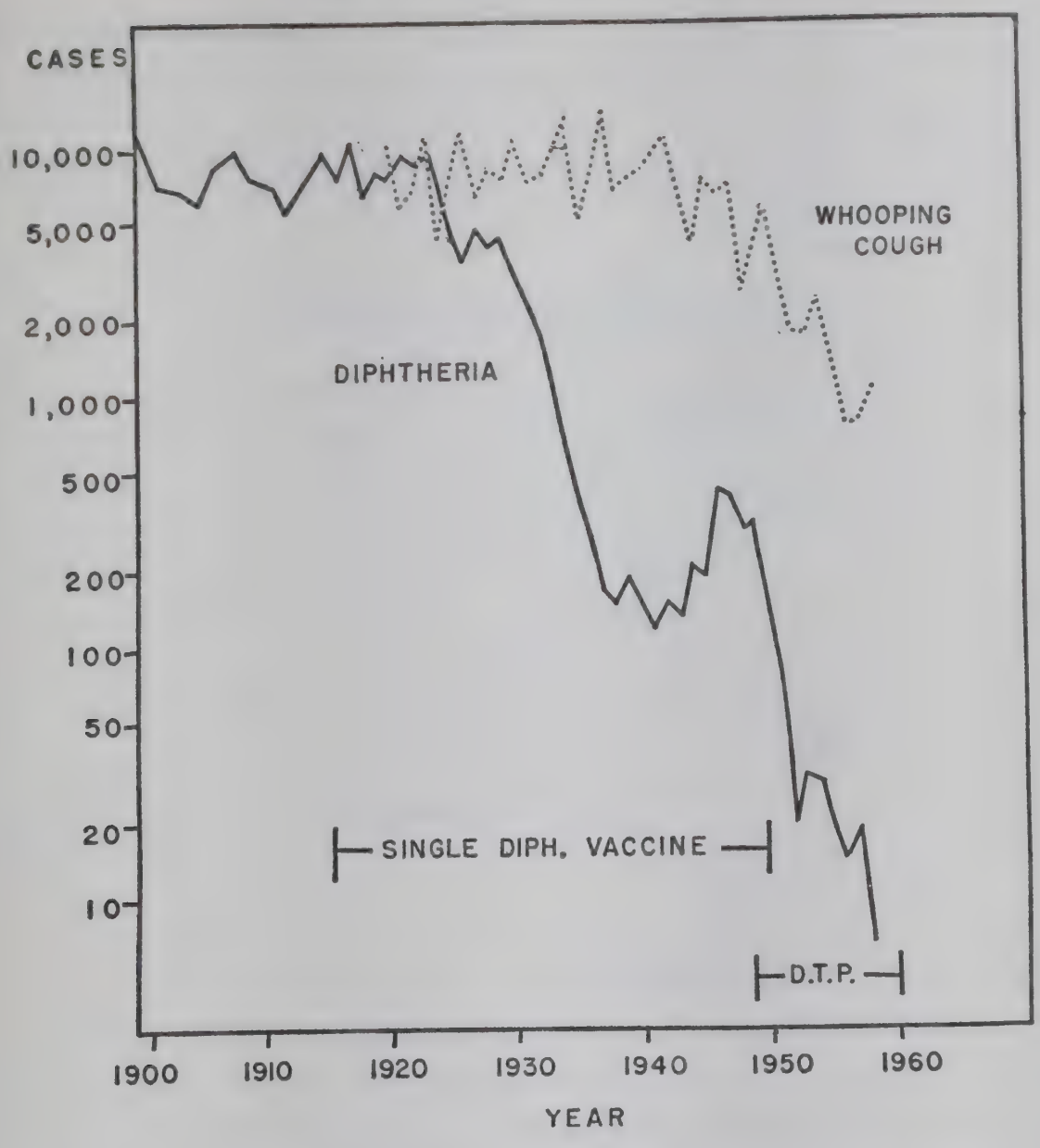


FIGURE 37. Reported cases of whooping cough and diphtheria in Massachusetts, 1909-1959. D.T.P.: diphtheria-tetanus-pertussis vaccine.

introduced with single antigens about 1920, while systematic whooping cough vaccination started in the late 1940's with the introduction of the combined diphtheria-tetanus-pertussis vaccine. Both diseases show striking decreases in frequency beginning about the time of the introduction of the respective specific measures. Patterns of this kind are often accepted as a valid demonstration of the effectiveness of a public health measure. It is interesting to note that diphtheria immunization was introduced without controlled experiments in human populations, while pertussis immunization was introduced only after rigorous field trials had demonstrated the effectiveness of the vaccine. The

downward trend in diphtheria incidence is therefore virtually the only evidence that exists for the effectiveness of diphtheria vaccine, while the downward trend of whooping cough incidence is an added confirmation of results obtained by more rigorous experimental methods.

CONTROLLED EXPERIMENT

Experimental epidemiology in the modern sense is a cohort study in which the difference between cohorts is in respect to some artificially induced characteristic. Where only one cohort has the induced characteristic, it is common to speak of the cohort with the characteristic as the "experimental group" and the one without as the "control group." A study may have several experimental groups, and the control group may be another form of experimental group in which the induced characteristic differs in kind or amount from that being tested. The essence of a well-designed trial is that the various experimental and control groups do not differ significantly from each other except with respect to the attribute under test.

Selection of Study Populations

For a field trial the selection of a population and the determination of its necessary size depend on prediction of incidence of disease. In a clinical trial prediction of prognosis is required. The expected incidence (of disease or of unfavorable outcome) together with estimates or guesses as to the differences that might be expected between the compared groups is the base for deciding on the size of the groups that will be required for the demonstration of statistically significant differences. Since predictions both of incidence and of differences between groups must of necessity be crude (otherwise the trial would not be required), it is usual to include considerably more persons in the study than statistical tests indicate to be the minimum required for confidence in the significance of predicted differences.

The more complete the epidemiologic knowledge of the determinants of the incidence and the course of disease, the more satisfactory the selection of population groups for inclusion is likely to be. A primary guide in the selection is the likelihood of ob-

serving a high incidence of disease in persons representative of the chosen population. Of no small importance, however, are additional factors such as the facilities available for observing such groups and the accessibility of individuals for detailed study. The latter considerations are of importance, for example, in selection of a geographic area suitable for field trials of measures directed against those infectious diseases that are predominant in countries with limited health resources. In spite of high frequency of the disease in one region, the lack of medical facilities is a consideration that may lead to preference for another where the expected incidence may be lower but where medical and public health resources permit precise diagnosis and adequate follow-up.

Study Design

An early field trial (1934 to 1937) serves as an example to illustrate some pitfalls in design that have led to demands for randomization of study groups and more rigorous conduct of observation. Kendrick and Eldering (1939) organized a trial of pertussis vaccine that, in spite of its theoretical drawbacks, elevated this prophylactic from a status of disputed and obscure significance to membership among the most useful of public health measures. The vaccine was offered on a voluntary basis to the preschool population of Grand Rapids, Michigan, through public immunization clinics. All children of that city were listed beforehand by several sociologic strata, and for each child that entered the vaccination roster a child of the same age and social stratum was selected for observation as an unvaccinated control. Table 42 shows attack rates among vaccinated and nonvaccinated children in three levels of exposure. As is evident from the difference in the totals of vaccinated and nonvaccinated children, the "matching" procedure was not exact.

The weaknesses in this study design are, first, that the subsequent observation of disease could be influenced by the knowledge of previous treatment and, second, that volunteering for treatment may be an attribute directly associated with determinants of general hygiene, disease awareness, and so on, which can affect the inception and outcome of disease.

The extensive field trials of pertussis vaccine conducted in England by the Medical Research Council from 1946 to 1950

TABLE 42

Incidence of whooping cough in a field trial of vaccines, Grand Rapids, Michigan, 1934-1937 *

Exposure group	Vaccinated			Not vaccinated		
	Total	Affected	Per cent	Total	Affected	Per cent
Exposed in own household	83	29	34.9	160	143	89.4
Exposed in other household	100	5	5.0	61	34	55.7
Indefinite exposure	114	4	3.5	52	10	19.2
No known exposure	1518	14	0.9	2124	161	7.6
Total	1815	52	2.9	2397	348	14.5

* From Kendrick and Eldering (1939).

(1951) were designed to eliminate these sources of uncertainty. Inclusion in this study was based on the free choice by the parents, but at the time of election to enter the study the participants were informed that the children would be given either pertussis vaccine or an ineffective placebo. The vaccine and placebo were packaged identically, distinguished only by code numbers unknown to the participating centers, and were distributed in a random fashion. Consequently, neither the participants nor the professional personnel conducting the trial knew which children had received the vaccine and which the placebo. Identification of the treatment remained in code until the follow-up observations were concluded. After the code was disclosed the randomization was found to have been almost complete with respect to numbers of children, distribution of home and other exposures, and other characteristics of the children. For all exposure categories attack rates were about five times higher in the control group than in the vaccinated group.

This study is an example of a method which is sometimes called the "double-blind" technique. This phrase implies (1) that the treatment is identified when administered only by a code, the meaning of which is unknown to those giving and receiving it, and (2) that assessment of outcome is also made without knowledge of the treatment given, or in fact of whether any treatment has been given. It is clear that either one or both of these objectives may be unattainable in particular situations.

Thus, the treatment under test may be one whose administration involves pain or considerable inconvenience to the patient and it may therefore be undesirable to impose similar inconvenience on persons who can derive no possible benefit. Even so, however, it may be possible to fulfill the second objective by having assessment of outcome made by persons other than those administering the treatment and under conditions in which they cannot identify the group to which the persons assessed belong. Where "blind" assessment is not possible, reliance must be placed on the selection of criteria of outcome that are as objective as possible.

Another important characteristic of the Medical Research Council vaccination study was the randomization of study groups. It is not necessary that the study groups be picked at random from the population. On the contrary, it is most useful to select groups according to determinants of expected incidence. The randomization is equally useful when it occurs *after* the decision has been made that this particular person should and will enter the study. The procedure of randomization must be prearranged with coded ampoules or sealed envelopes containing the assigned treatment. Other methods, such as assigning on the basis of date of admissions, first letter in names, and so on, are undesirable because the treated persons are identifiable and unconscious manipulations are possible.

Like the "double-blind" technique, randomization is also not always practical. For example, in testing the effect of addition of fluoride to drinking water on the frequency of dental caries the practical problems of administration of the treatment require that the treatment be either given or not given to entire populations having single sources of drinking water, with no possibility existing of random allocation within those populations. Since many persons will necessarily be involved in such an experiment, it will also be necessary that the experiment be limited to two or at most very few populations. Validity of results will then depend on (1) careful selection of experimental and control populations so that they are as similar as possible in all conceivably relevant characteristics, (2) detailed comparison of experimental and control populations with respect to the outcome to be assessed (e.g., dental caries) *before* introduction of the treatment, and (3) assessment of outcome by observers who are unaware of the

population to which the assessed persons belong. These criteria are fulfilled and well illustrated by the fluoridation experiments undertaken by the U. S. Public Health Service (Ast *et al.,* 1956; Arnold, 1957).

Sequential Designs

In a classic "double-blind" experiment, the code may not be disclosed before observations are complete on a group whose size was determined at the outset of the experiment, and sometimes on the basis of incomplete information. The investigators thereby expose themselves to the dilemma that even though significant results may have been obtained halfway through the experiment, because of superior effectiveness of the treatment under test, they will not have known about it. If this is the case, an undue number of "control" persons will have been jeopardized by not receiving the treatment. This dilemma mostly pertains to clinical trials where patients are admitted over a long period. The problem of an untoward effect of the treatment also exists; the investigators may be faced with the risk that the treatment is not only ineffective but even noxious. To safeguard against this it is often suggested in clinical trials that the code be revealed to the physician in charge in the event of medical emergencies, because of his professional right to know all facts about his patient. Perhaps this clause in the design is used much less frequently than one would assume, since the untoward effect may not show up dramatically in one patient as much as by group comparison. For such reasons, the current tendency is to develop designs for clinical trials based on the principle of sequential design. This means that the trial is interrupted as soon as statistical significance has been obtained in one direction or the other. There is no predetermined group size. This does not necessarily conflict with the "double-blind" principle because the decision to stop the trial for statistical reasons can be made by a member of the study group who is not participating in the clinical work but makes his decision on the basis of results submitted to him sequentially.

Another design, sometimes called the semi-sequential design, requires that the code be disclosed and a new one substituted each time the study of a predetermined small number of subjects is con-

cluded. This number may, for instance, be 20 subjects, so that the sequential analyses are made on consecutive sets of 20 and the decision to interrupt is made when statistical significance is obtained.

Multiple Field Trials

Field trials are self-limiting by nature. If a well-conducted trial shows the beneficial effect of a treatment or prophylactic measure, ethics demand that further trials cannot be conducted with untreated controls and any future demonstration of the principle involved in the treatment must consist of a comparison between the treatment established in the first field trial and suggested improvements over this treatment. One could speak of the phase 1 trial where the "control" is no treatment; such a trial is justified when the efficacy of treatment is unknown. A phase 2 trial would be one where treatment in some form is known to be effective but where other forms of the same treatment might prove more effective. The Medical Research Council (1956) conducted a second trial of pertussis vaccine based on this principle. The trial, which took place from 1951 to 1954, was concerned with the relative efficacy of fourteen vaccines of different manufacture. All of the 19,000 participating children were vaccinated with a pertussis vaccine but the field trial showed significant differences in the attack rates among children inoculated with the different vaccines. These attack rates correlated conclusively with the relative potencies of the individual vaccines as measured in assays in laboratory animals.

Trials and Practice

Because of the limitation on the number of factors that can be varied in costly field trials, the interpretation of results in terms of general applicability may be limited. In addition there may be an inability to translate a method successfully applied in a field trial into a prescription for practical use. An outstanding example of the latter was consequent on the field trial of gamma-globulin for the prevention of poliomyelitis conducted by Hammon *et al.* (1952). This well-planned trial demonstrated that injection of human gamma-globulin led to a significantly lower incidence of paralytic poliomyelitis than that experienced by placebo-injected controls if the injections were close to time of known exposure to

infection. The interest of the trial was primarily to test the hypothesis that specific antibodies are important determinants in the inception of paralytic poliomyelitis. As a result of that finding hundreds of thousands of children in the United States were injected with gamma-globulin during the 1953 poliomyelitis season, but without any demonstration of resulting decrease in incidence of the disease. Presumably the lack of effect was due to an inability to apply generally the rigid criteria that were used in the trial.

In a field trial of poliomyelitis vaccine in 1954 (Francis *et al.*, 1957), school children aged 5 to 9 years were selected solely because expected incidence was highest in that age group and because of the facility of enlisting school children in an epidemiologic study. The successful results of the trial led to immediate use of the vaccine, but it is quite conceivable that the focus on school children in the trial led to more extensive use of the vaccine in this age group than in the preschool child. At least, the situation five years later, in 1959, showed that the children of school age were immunized in a significantly higher proportion than the preschool children. This fact was reflected in an attack rate almost three times higher in the preschool children than in the younger school children in 1959.

"NATURAL" EXPERIMENTS

From time to time, circumstances arise without human intent that have many of the features of experiment. Although lacking the human intervention that makes an experiment an experiment, these circumstances are frequently referred to as "natural experiments."

An example of a "natural" experiment reported by Hyge (1947) concerns an outbreak of pulmonary tuberculosis in a day school for girls. The epidemic took place 2 months after the whole student body of 368 had been examined by tuberculin test and chest X-ray. About a third of the students had a negative tuberculin test, another third had a positive test from natural sources, and the remainder were positive after BCG vaccination conducted 8 months previously. The outbreak was indisputably traced to a science teacher with a pulmonary lesion that was active intermittently. This teacher taught eight different grades

comprising 212 students of all ages between 10 and 18 years. Four other classes totaling 120 students used the same classroom in which she had taught, but after she had conducted classes. Finally, there were three classes (36 girls) who had none of these exposures to the teacher. Table 43, showing the cases of pulmonary tuberculosis that developed during the next five years, illustrates a design that could scarcely have been improved if 368 guinea pigs had been distributed in three equal groups by vaccination status and each with three levels of intensity of exposure. The effect of the BCG vaccination is striking in these data.

However, it must be borne in mind that the study would hardly have been published but for the unusual nature of the circumstances that comprised this event. Outstanding among these is the extreme unlikelihood that a tuberculosis incidence of this magnitude would occur such a short time after a whole population had been tested for presence of infection. The very rarity of such events stamps them as unusual and brings into doubt the validity of generalizations based on the experience.

Evidence derived from "natural" experiments rarely is convincing, and for evidence to be as convincing as that given in this example is extremely uncommon. The major concern is, of course, the lack of randomization in the selection of the groups. It is extremely difficult to rule out the possibility of serious bias in such selection.

However, in comparison with field trials, the observation of such events is inexpensive, and eliminates the necessity of considering ethical aspects inherent in experimentation. It would be against good scientific judgment to deny the usefulness of such observations.

CAUTION

Perhaps the greatest advance in medicine and public health in the last decade has been the acceptance of the role of scientifically designed trials in the testing of hypotheses. In the light of the many practical considerations, none of these trials may ever reach theoretical perfection, but it is no exaggeration to say that the acceptance of the principles involved indicates a revolution in medical thinking.

TABLE 43

Frequency of gross tuberculous disease among girls in a Danish school during a 5 year period following exposure to a known source of infection: by degree of exposure and status before exposure†*

Degree of exposure	Status before exposure									Total		
	Mantoux positive						Mantoux negative					
	No BCG			BCG vaccinated								
	Total	Affected	Per cent	Total	Affected	Per cent	Total	Affected	Per cent	Total	Affected	Per cent
Taught by X	74	1	1.4	86	2	2.3	52	16	30.8	212	19	9.0
Taught in X's room but not by X	40	1	2.5	32	0	0	48	5	10.4	120	6	5.0
Neither	16	0	0	15	0	0	5	0	0	36	0	0
Total	130	2	1.5	133	2	1.5	105	21	20.0	368	25	6.8

* Progressive pulmonary tuberculosis (14 patients); pleurisy, pericaditis or peritonitis (11 patients).
† From Hyge (1947, 1949).

However, care is required to insure that the respect due the epidemiologic experiment does not swell to the extent that observational methods become ignored or disfavored. At times, even now, the results of sound observation are denied because experimental "proof" is lacking. We have pointed out previously that there are many situations where experiment is impossible and many more where it is impractical. Such situations are abundant in the areas with which epidemiology is now particularly concerning itself — the chronic diseases — and it is likely that, as has happened so frequently in the past, many effective control measures will be instituted without experimental confirmation of hypotheses based on observation.

Bibliography

Allison, A. C. 1954. The distribution of the sickle-cell trait in East Africa and elsewhere, and its apparent relationship to the incidence of subtertian malaria. *Tr. Roy. Soc. Trop. Med. & Hyg. 48:* 312.

American Public Health Association 1960. *The Control of Communicable Diseases in Man* (9th ed.). American Public Health Association, New York.

Andvord, K. F. 1930. What can we learn by studying tuberculosis by generations? *Norsk mag. laegevidensk., 91:* 642.

Arnold, F. A., Jr. 1957. Grand Rapids fluoridation study — results pertaining to the eleventh year of fluoridation. *Am. J. Pub. Health, 47:* 539.

Ast, D. B., Smith, D. J., Wachs, B., and Cantwell, K. T. 1956. Newburgh-Kingston caries-fluorine study. XIV. Combined clinical and roentgenographic dental findings after ten years of fluoride experience. *J. Am. Dent. A., 52:* 314.

Aycock, W. L., and Luther, E. H. 1929. The occurrence of poliomyelitis following tonsillectomy. *New England J. Med., 200:* 164.

Bailey, A. H., and Day, A. 1861. On the rate of mortality prevailing among the families of the peerage during the nineteenth century. *Assurance Mag. and J. Inst. Actuaries, 9:* 4.

Baker, G. 1767. *An Essay Concerning the Cause of the Endemial Colic of Devonshire*. Reprinted by Delta Omega Society, American Public Health Association, New York, 1958.

Bancroft, H. 1957. *Introduction to Biostatistics.* Hoeber-Harper, New York.

Beeton, M., and Pearson, K. 1901. On the inheritance of the duration of life, and on the intensity of natural selection in man. *Biometrika, 1:* 50.

Berman, C. 1940. Primary carcinoma of the liver in the Bantu races of South Africa. *South African J. M. Sc., 5:* 54.

Cairns, M., and Stewart, A. 1951. Pulmonary tuberculosis mortality in the printing and shoemaking trades: historical survey, 1881-1931. *Brit. J. Social Med., 5:* 73.

Carter, C., and MacCarthy, D. 1951. Incidence of mongolism and its diagnosis in the newborn. *Brit. J. Social Med., 5:* 83.

Case, R. A. M. 1956. Cohort analysis of mortality rates as an historical or narrative technique. *Brit. J. Prev. & Social Med., 10:* 159.

Case, R. A. M., Hosker, M. E., McDonald, D. B., and Pearson, J. T. 1954. Tumors of the urinary bladder in workmen engaged in the manufacture and use of certain dyestuff intermediates in the British chemical industry. *Brit. J. Indust. Med., 11:* 75.

Case, R. A. M., and Lea, A. J. 1955. Mustard gas poisoning, chronic bronchitis, and lung cancer. *Brit. J. Prev. & Social Med., 9:* 62.

Case, R. A. M., and Pearson, J. T. 1956. *Cancer Death Rates by Site, Age and Sex; England and Wales, 1911-55.* Institute of Cancer Research, Royal Cancer Hospital, London (Mimeographed).

Caverly, C. S. 1894. Preliminary report of an epidemic of paralytic disease, occurring in Vermont, in the summer of 1894. *Yale Med. J., 1:* 1.

Chapin, C. V. 1924. Deaths among taxpayers and non-taxpayers of income tax, Providence, 1865. *Am. J. Pub. Health, 14:* 647.

Ciocco, A. 1940. On the mortality in husbands and wives. *Human Biol., 12:* 508.

Clemmesen, J., and Nielsen, A. 1957. Comparison of age-adjusted cancer incidence rates in Denmark and the United States. *J. Nat. Cancer Inst., 19:* 989.

Cochran, W. G. 1951. Modern methods in sampling of human populations: general principles in selection of sample. *Am. J. Pub. Health, 41:* 647.

Collins, S. D. 1933. Causes of illness in 9,000 families based on nation-wide periodic canvasses, 1928-1931. *Pub. Health Rep., 48:* 283.

Collins, S. D. 1940. Cases and days of illness among males and females, with special reference to confinement to bed. *Pub. Health Rep., 55:* 47.

Commission on Chronic Illness 1957. *Chronic Illness in the United States. Volume IV. Chronic Illness in a Large City: The Baltimore Study.* Harvard University Press, Cambridge, Massachusetts.

Commission on Chronic Illness 1959. *Chronic Illness in the United States. Volume III. Chronic Illness in a Rural Area: The Hunterdon Study.* Harvard University Press, Cambridge, Massachusetts.

Communicable Disease Center 1957. *Influenza Surveillance Report,* Communicable Disease Center, U. S. Public Health Service, Atlanta, Georgia (weekly).

Court-Brown, W. M., and Doll, R. 1957. *Leukemia and Aplastic Anaemia in Patients Irradiated for Ankylosing Spondylitis.* Medical Research Council Special Report Series No. 295. Her Majesty's Stationery Office, London.

Croxton, F. E. 1953. *Elementary Statistics with Applications in Medicine.* Prentice-Hall, Inc., New York.

Cullen, W. 1785. *Synopsis Nosologicae Medicae.* Quoted by Hosack, D. *A System of Practical Nosology* (2nd ed.). O. S. Van Winkle, New York, 1821.

Dawber, T. R., Meadors, G. F., and Moore, F. E., Jr. 1951. Epidemiological approaches to heart disease. The Framingham Study. *Am. J. Pub. Health, 41:* 279.

Dean, H. T. 1946. Some general epidemiological considerations. Chapter I in *Dental Caries and Fluorine.* Ed., Moulton, F. R. *Am. Assoc. Adv. Sc.,* Washington, D. C.

Doll, R., and Hill, A. B. 1952. A study of the aetiology of carcinoma of the lung. *Brit. M. J., 2:* 1271.

Doll, R., and Hill, A. B. 1954. The mortality of doctors in relation to their smoking habits; a preliminary report. *Brit. M. J., 1:* 1451.

Doll, R., and Hill, A. B. 1956. Lung cancer and other causes of death in relation to smoking. A second report on the mortality of British doctors. *Brit. M. J., 2:* 1071.

Dorn, H. F. 1950. Methods of analysis for follow-up studies. *Human Biol., 22:* 238.

Dorn, H. F. 1959a. Mortality. Chapter 19 in *The Study of Population.* Ed., Hauser, P. M., and Duncan, O. D. University of Chicago Press, Chicago, Illinois.

Dorn, H. F. 1959b. Tobacco consumption and mortality from cancer and other diseases. *Pub. Health Rep., 74:* 581.

Dorn, H. F., and Cutler, S. J. 1959. *Morbidity from Cancer in the United States.* Public Health Mono. No. 56, U. S. Govt. Printing Office, Washington, D. C.

Downes, J. 1947. Chronic disease among spouses. *Milbank Mem. Fund Quart., 25:* 334.

Downes, J. 1950. Cause of illness among males and females. *Milbank Mem. Fund Quart., 28:* 407.

Downes, J., and Collins, S. D. 1940. A study of illness among families in the Eastern Health District of Baltimore. *Milbank Mem. Fund Quart., 18:* 5.

Doyle, J. T., Heslin, A. S., Hilleboe, H. E., Formel, P. F., and Korns, R. F. 1957. A prospective study of degenerative cardiovascular disease in Albany: report of three years' experience — 1. Ischemic heart disease. *Am. J. Pub. Health,* Suppl., April, 1957, p. 25.

Dublin, L. I., and Marks, H. H. 1941. Inheritance of longevity — study based on life insurance records. *Tr. Ass. Life Insur. Med. Dir. America, 28:* 89.

Duncan, J. M. 1865. On some laws of the production of twins. *Edinburgh M. J., 10:* 767.

Dunn, H. L., and Shackley, W. 1944. Comparison of cause-of-death assignments by the 1929 and 1938 revisions of the International List: deaths in the United States, 1940. *Vital Statistics Special Reports, 19:* 14.

Dunning, J. M. 1953. The influence of latitude and distance from seacoast on dental disease. *J. Dent. Res., 32:* 811.

Eastcott, D. F. 1956. The epidemiology of lung cancer in New Zealand. *Lancet, 1:* 37.

Eaton, J. W., and Weil, R. J. 1955. *Culture and Mental Disorders.* The Free Press, Glencoe, Illinois.

Enos, W. F., Jr., Beyer, J. C., and Holmes, R. H. 1955. Pathogenesis of coronary disease in American soldiers killed in Korea. *J. A. M. A., 158:* 912.

Erhardt, C. L., and Weiner, L. 1950. Changes in mortality statistics through use of the new international statistical classification. *Am. J. Pub. Health, 40:* 6.

Fogh-Andersen, P. 1942. *Inheritance of Hare-lip and Cleft Palate.* Busck, Copenhagen.

Francis, T., Jr., Napier, J. A., Voight, R. B., Hemphill, F. M., Wenner, H. A., Korns, R. S., Boisen, M., Tolchinsky, E., and Diamond, E. L. 1957. *Evaluation of the 1954 Field Trial of Poliomyelitis Vaccine. Final report.* Poliomyelitis Vaccine Evaluation Center, University of Michigan, Ann Arbor, Michigan.

Fraser, R., Bunbury, E., Daniell, B., Barling, M. E., Waldron, F. E., Kemp,

P. M., and Lee, I. 1947. *The Incidence of Neurosis Among Factory Workers.* Reports of the Industrial Health Research Board, No. 90., Her Majesty's Stationery Office, London.

Friedman, M., Rosenman, R. H., and Carroll, V. 1958. Changes in the serum cholesterol and blood clotting time in men subjected to cyclic variation of occupational stress. *Circulation, 17:* 852.

Frost, W. H. 1939. The age selection of mortality from tuberculosis in successive decades. *Am. J. Hyg.* Sect. A, *30:* 91.

Galton, F. 1892. *Hereditary Genius.* Macmillan, London.

Gentry, J. T., Nitowsky, H. M., and Michael, M., Jr. 1955. Studies on the epidemiology of sarcoidosis in the United States, relationship to soil areas and to urban-rural residence. *J. Clin. Invest., 34:* 1839.

Gentry, J. T., Parkhurst, E., and Bulin, G. V., Jr. 1959. An epidemiological study of congenital malformations in New York State. *Am. J. Pub. Health, 49:* 497.

Gilliam, A. G. 1954. A note on evidence relating to the incidence of primary liver cancer among the Bantu. *J. Nat. Cancer Inst., 15:* 195.

Gilliam, A. G. 1955. Trends of mortality attributed to carcinoma of the lung: possible effects of faulty certification of deaths to other respiratory diseases. *Cancer, 8:* 1130.

Gilliam, A. G. 1960. National Cancer Institute, U. S. Public Health Service. Unpublished data.

Goldberger, J., Wheeler, G. A., and Sydenstricker, E. 1920. A study of the relation of family income and other economic factors to pellagra incidence in seven cotton-mill villages of South Carolina in 1916. *Pub. Health Rep., 35:* 2673.

Greenwood, M., Hill, A. B., Topley, W. W. C., and Wilson, J. 1936. *Experimental Epidemiology.* M.R.C. Special Report Series No. 209. His Majesty's Stationery Office, London.

Greenwood, M., and Yule, G. U. 1914. On the determination of size of family and of the distribution of characters in order of birth from samples taken through members of the sibships. *J. Roy. Statist. Soc., 77:* 179.

Gregg, N. M. 1942. Congenital cataract following German measles in the mother. *Tr. Ophth. Soc.* Australia, 1941, *3:* 35.

Griswold, M. H., Wilder, C. S., Cutler, S. J., and Pollack, E. S. 1955. *Cancer in Connecticut 1935-1951.* Connecticut State Department of Health, Hartford, Conn.

Hadley, J. N. 1955. Health conditions among Navajo Indians. *Pub. Health Rep., 70:* 831.

Haenszel, W., Marcus, S. C., and Zimmerer, E. G. 1956. *Cancer Morbidity in Urban and Rural Iowa.* Public Health Mono. No. 37, Public Health Service Publication No. 462. U. S. Govt. Printing Office, Washington, D. C.

Hammon, W. McD., Coriell, L. L. and Stokes, J., Jr. 1952. Evaluation of Red Cross gamma globulin as a prophylactic agent for poliomyelitis. *J. A. M. A., 150:* 739.

Hammond, E. C., and Horn, D. 1958. Smoking and death rates — report on 44 months of follow-up of 187,783 men. *J. A. M. A., 166:* 1159 and 1294.

Hempel, C. G. 1959. Some problems of taxonomy. Work conference on problems in field studies in the mental disorders. *Amer. Psychopath. Assoc.,* New York.

Hill, A. B. 1948. *Principles of Medical Statistics* (4th ed.). The Lancet Ltd., London.

Hill, A. B., Doll, R., Galloway, T. M., and Hughes, J. P. W. 1958. Virus diseases in pregnancy and congenital defects. *Brit. J. Prev. & Social Med., 12:* 1.

Hirsch, A. 1883-1886. *Handbook of Historical and Geographical Pathology. Vols. I-III.* Translated from the Second German Edition by Creighton, C. The New Sydenham Society, London.

Hollingshead, A. B., and Redlich, F. C. 1958. *Social Class and Mental Illness: A Community Study.* John Wiley & Sons, Inc., New York.

Holt, S. B. 1948. The effect of maternal age on the manifestation of a polydactyl gene in mice. *Ann. Eugenics, 14:* 144.

Hueper, W. C. 1950. *A Methodology for Environmental and Occupational Cancer Surveys.* Public Health Technical Mono. No. 1 U. S. Govt. Printing Office, Washington, D. C.

Hueper, W. C. 1952. Age aspects of environmental and occupational cancers. *Pub. Health Rep., 67:* 773.

Hueper, W. C. 1954. Recent developments in environmental cancer. *Arch. Path., 58:* 360, 475 and 645.

Hume, D. 1739. *Treatise of Human Nature.* Ed., Selby-Bigge, Oxford, Clarendon, 1896.

Huxley, J. 1951. Genetics, evolution and human destiny. In *Genetics in the 20th Century.* Ed., Dunn, L. C. Macmillan, New York.

Hyge, T. V. 1947. Epidemic of tuberculosis in a state school, with observation period of about 3 years. *Acta tuberc. scandinav., 21:* 1.

Hyge, T. V. 1949. The efficacy of BCG vaccination. *Acta tuberc. scandinav., 22:* 153.

Ipsen, J., Jr., 1950. Prevalence and incidence of multiple sclerosis in Boston 1939-1948. *Arch. Neurol. & Psychiat., 64:* 631.

Jalavisto, E. 1950. The influence of parental age on the expectation of life. *Rev. méd. Liége, 5:* 719.

Kendrick, P. and Eldering, G. 1939. A study in active immunization against pertussis. *Am. J. Hyg., 29:* (Sec. B) 133.

Kettle, E. H. 1934. Experimental pneumoconiosis: infective silicatosis. *J. Path. & Bact., 38:* 201.

Keys, A., Fidanza, F., Scardi, V., Bergami, G., Keys, M. H. and Dilorenzo, F. 1954. Studies on serum cholesterol and other characteristics of clinically healthy men in Naples. *Arch. Int. Med., 93:* 328.

Knobloch, H., and Pasamanick, B. 1958. Seasonal variation in the births of the mentally deficient. *Am. J. Pub. Health, 48:* 1201.

Koch, R. 1882. The aetiology of tuberculosis. Berlin. *Klin. Wchnschr., 19:* 221. Translated and reprinted in *The Aetiology of Tuberculosis,* National Tuberculosis Assoc., New York, 1932.

Korteweg, R. 1951. The age curve in lung cancer. *Brit. J. Cancer, 5:* 21.

Kraepelin, E. 1913. *General Paresis.* Authorized English Translation by Moore, J. W., Nervous and Mental Disease Monograph Series, No. 14. J. Nerv. and Ment. Dis. Publishing Co., New York.

Kutschenreuter, P. H. 1959. A study of the effect of weather on mortality. Paper given at a meeting of the Division of Oceanography and Meteorology, New York Academy of Sciences, Nov. 24th.

Landauer, W., and Beall, G. 1952. Selection and the age incidence of human cancer. *Am. J. Hum. Genet., 4:* 173.

Lange, R. D., Moloney, W. C., and Yamawaki, T. 1954. Leukemia in atomic bomb survivors. I. General considerations. *Blood, 9:* 574.

Lansing, A. I. 1947. A transmissible, cumulative, and reversible factor in aging. *J. Gerontol., 2:* 228.

Lansing, A. I. 1954. A nongenic factor in the longevity of rotifers. *Ann. New York Acad. Sc., 57:* 455.

Lee, R. E., and Schneider, R. F. 1958. Hypertension and arteriosclerosis in executive and nonexecutive personnel. *J. A. M. A., 167:* 1447.

Leighton, A. 1959. *My Name is Legion.* Basic Books, New York.

Lilienfeld, A. M. 1956. The relationship of cancer of the female breast to artificial menopause and marital status. *Cancer, 9:* 927.

Linder, F. E. 1958. National Health Survey. *Science, 127:* 1275.

Livingstone, F. B. 1958. Anthropological implications of sickle cell gene distribution in West Africa. *Am. Anthrop., 60:* 533.

Logan, W. P. D. 1953. Marriage and childbearing in relation to cancer of the breast and uterus. *Lancet, 2:* 1199.

Lowe, C. R. 1956. An association between smoking and respiratory tuberculosis. *Brit. M. J., 2:* 1081.

Lowe, C. R., and Garratt, F. N. 1959. Sex pattern of admissions to mental hospitals in relation to social circumstances. *Brit. J. Prev. & Social Med., 13:* 88.

MacMahon, B. 1957. Epidemiologic evidence on the nature of Hodgkin's disease. *Cancer, 10:* 1045.

MacMahon, B., and Feinleib, M. 1960. Breast cancer in relation to nursing and menopausal history. *J. Nat. Cancer Inst., 24:* 733.

MacMahon, B., Pugh, T. F., and Ingalls, T. H. 1953. Anencephalus, spina bifida, and hydrocephalus: incidence related to sex, race, and season of birth, and incidence in siblings. *Brit. J. Prev. & Social Med., 7:* 211.

MacMahon, B., Record, R. G., and McKeown, T. 1951. Congenital pyloric stenosis — An investigation of 578 cases. *Brit. J. Social Med., 5:* 185.

MacMahon, B., and Terry, W. D. 1958. Application of cohort analysis to the study of time trends in neoplastic disease. *J. Chronic Dis., 7:* 24.

Mainland, D. 1954. Chance and random sampling. *Methods in Med. Res., 6:* 127.

Malzberg, B. 1940. *Social and Biological Aspects of Mental Disease.* State Hospitals Press, Utica, New York.

Malzberg, B., and Lee, E. S. 1956. *Migration and Mental Disease, A Study of First Admissions to Hospitals for Mental Disease, New York, 1939-41.* Social Science Research Council, New York.

Manos, N. E. 1957. *Comparative Mortality Among Metropolitan Areas of the United States, 1949-51; 102 Causes of Death.* Public Health Service Publication No. 562, U. S. Govt. Printing Office, Washington, D. C.

Mantel, N., and Haenszel, W. 1959. Statistical aspects of the analysis of data from retrospective studies of disease. *J. Nat. Cancer Inst., 22:* 719.

Martin, W. J. 1955. Vital statistics of the county of London in the years 1901 to 1951. *Brit. J. Prev. & Social Med., 9:* 126.

Masland, R. L. 1958. A collaborative project for the study of cerebral palsy, mental retardation and other neurological and sensory disorders of child-

hood. Presented at the 86th Annual Meeting, Amer. Public Health Association, St. Louis, October 27 (Mimeographed).

McDonald, A. D. 1958. Maternal health and congenital defect: a prospective investigation. *New England J. Med., 258:* 767.

McKeown, T., MacMahon, B., and Record, R. G. 1951. The incidence of congenital pyloric stenosis related to birth rank and maternal age. *Ann. Eugenics, 16:* 249.

McKeown, T., MacMahon, B., and Record, R. G. 1955. Evidence of postnatal environmental influence in the aetiology of infantile pyloric stenosis. *Arch. Dis. Childhood, 27:* 386.

McKeown, T., and Record, R. G. 1956. Maternal age and birth order as indices of environmental influence. *Amer. J. Human Genet., 8:* 8.

Medical Research Council 1951. The prevention of whooping cough by vaccination. *Brit. M. J., 1:* 1463.

Medical Research Council 1956. Vaccination against whooping-cough. *Brit. M. J., 2:* 454.

Menaker, W. 1956. Air pollution, smoking, and lung cancer (letter to the editor). *Brit. M. J., 2:* 416.

Mill, J. S. 1862. *A System of Logic, Ratiocinative and Inductive* (5th ed.). Parker, Son and Bowin, London.

Miller, D. C., Trulson, M. F., McCann, M. B., White, P. D., and Stare, F. J. 1958. Diet, blood lipids and health of Italian men in Boston. *Ann. Int. Med., 49:* 1178.

Minister of Health 1954. *Mortality and Morbidity During the London Fog of December, 1952.* Report by a Committee of Departmental Officers and Expert Advisers appointed by the Minister of Health. Reports on Public Health and Medical Subjects No. 95. Her Majesty's Stationery Office, London.

Morris, J. N., Heady, J. A., Raffle, P. A. B., Roberts, C. G., and Parks, J. W. 1953. Coronary heart-disease and physical activity of work. *Lancet, 2:* 1053.

Muench, H. 1959. *Catalytic Models in Epidemiology.* Harvard University Press, Cambridge, Massachusetts.

National Cancer Institute 1960. Cooperative study of childhood leukemia. In progress.

National Health Survey 1958. *Origin and Program of the U. S. National Health Survey.* Public Health Service Publ. No. 584-A1. U. S. Govt. Printing Office, Washington, D. C.

National Health Survey 1959. *Disability Days, United States, July 1957—June 1958.* Public Health Service Publication No. 584-B10. Superintendent of Documents, Washington, D. C.

National Office of Vital Statistics Annual. *Vital Statistics of the United States.* Public Health Service, U. S. Govt. Printing Office, Washington, D. C.

National Office of Vital Statistics 1945. Deaths from selected causes by marital status, by age and sex: United States, 1940. *Vital Statistics Special Reports, 23:* No. 7, 117.

National Office of Vital Statistics 1951. *Vital Statistics Special Reports—National Summaries, 36:* 10.

National Office of Vital Statistics 1956. Mortality from selected causes by mari-

tal status, United States, 1949-51. *Vital Statistics Special Reports, 39:* No. 7, 303.

National Office of Vital Statistics 1957. Provisional information on selected notifiable diseases in the United States and on deaths in selected cities for week ended December 21, 1957. *Morbidity and Mortality Weekly Report,* Dec. 28, 1957.

Norton, A. 1952. Incidence of neurosis related to maternal age and birth order. *Brit. J. Social Med., 6:* 253.

Novitski, E., and Sandler, L. 1956. The relationship between parental age, birth order and the secondary sex ratio in humans. *Ann. Human Genet., 21:* 123.

Opler, M. K. 1958. Epidemiological studies of mental illness. Methods and scope of the Midtown study in New York. In *Symposium on Preventive and Social Psychiatry,* Walter Reed Army Institute of Research, Washington, D. C.

Panum, P. L. 1846. Observations made during the epidemic of measles on the Faroe Islands in the year 1846. Reproduced in *Panum on Measles,* Delta Omega Society. American Public Health Association, New York, 1940.

Penrose, L. S. 1934. A method of separating the relative aetiological effects of birth order and maternal age, with special reference to mongolian imbecility. *Ann. Eugenics, 6:* 108.

Perrott, G. St.J., Tibbitts, C., and Britten, R. H. 1939. The National Health Survey — Scope and method of the Nation-wide canvass of sickness in relation to its social and economic setting. *Pub. Health Rep., 54:* 1663.

Pintner, R., and Forlano, G. 1943. Season of birth and mental differences. *Psychol. Bull., 40:* 25.

Pollard, A. H. 1949. Methods of forecasting mortality using Australian data. *J. Institute of Actuaries, 75:* 151.

Pope, A. S., and Gordon, J. E. 1955. The impact of tuberculosis on human populations. *Am. J. M. Sc., 230:* 317.

Record, R. G., and McKeown, T. 1949. Congenital malformations of the central nervous system. I. A survey of 930 cases. *Brit. J. Social Med., 3:* 183.

Reed, S. C. 1936. Harelip in the house mouse. I. Effects of the external and internal environments. *Genetics, 21:* 339.

Registrar General for England and Wales Annual. *The Registrar General's Statistical Review of England and Wales.* Tables, Part I (Medical), Part II (Civil). Her Majesty's Stationery Office, London.

Registrar General for England and Wales 1938. *The Registrar General's Decennial Supplement, England and Wales, 1931. Part IIa. Occupational Mortality.* His Majesty's Stationery Office, London.

Registrar General for England and Wales 1947. *The Registrar General's Statistical Review of England and Wales for the Years 1938 and 1939.* Her Majesty's Stationery Office, London, Text. p. 99.

Registrar General for England and Wales 1954. *The Registrar General's Decennial Supplement, England and Wales, 1951. Occupational Mortality, Part I.* Her Majesty's Stationery Office, London.

Registrar General for England and Wales 1955. *The Registrar General's Statistical Review of England and Wales for the Two Years 1950-1951.* Supplement on General Morbidity, Cancer and Mental Health. Her Majesty's Stationery Office, London.

Sawyer, W. A., Meyer, K. F., Eaton, M. D., Bauer, J. H., Putnam, P., and Schwentker, F. F. 1944. Jaundice in army personnel in the Western Region of the United States and its relation to vaccination against yellow fever. *Am. J. Hyg., 39:* 337.

Scrimshaw, N. S., Trulson, M., Tejada, C., Hegsted, D. M., and Stare, F. J. 1957. Serum lipoprotein and cholesterol concentrations: comparison of rural Costa Rican, Guatemalan, and United States populations. *Circulation, 15:* 805.

Seltser, R., and Sartwell, P. E. 1959. The application of cohort analysis to the study of ionizing radiation and longevity of physicians. *Am. J. Pub. Health. 49:* 1610.

Shope, R. E. 1931. Swine influenza. III. Filtration experiments and etiology. *J. Exper. Med., 54:* 373.

Simpson, R. E. H. 1952. Infectiousness of communicable diseases in the household (measles, chickenpox, and mumps). *Lancet, 2:* 549.

Simpson, C. L., Hempelmann, L. H., and Fuller, L. M. 1955. Neoplasia in children treated with X-rays in infancy for thymic enlargement. *Radiology, 64:* 840.

Slater, P. 1946. *Survey of Sickness, October 1943 to December 1945.* The Social Survey. Her Majesty's Stationery Office, London.

Smith, R. L. 1956a. Recorded and expected mortality among the Japanese of the United States and Hawaii, with special reference to cancer. *J. Nat. Cancer Inst., 17:* 459.

Smith, R. L. 1956b. Recorded and expected mortality among the Chinese of Hawaii and the United States, with special reference to cancer. *J. Nat. Cancer Inst., 17:* 667.

Smith, R. L., Salsbury, C. G., and Gilliam, A. G. 1956. Recorded and expected mortality among the Navajo, with special reference to cancer. *J. Nat. Cancer Inst., 17:* 77.

Snow, J. 1855. *On the Mode of Communication of Cholera* (2nd ed.). John Churchill, London, Reproduced in *Snow on Cholera,* The Commonwealth Fund, New York, 1936.

Springett, V. H. 1952. An interpretation of statistical trends in tuberculosis. *Lancet, 1:* 521 and 575.

Stewart, A., Webb, J., and Hewitt, D. 1958. A survey of childhood malignancies. *Brit. M. J., 1:* 1495.

Still, G. F. 1927. Place-in-family as a factor in disease. *Lancet, 2:* 795 and 853.

Stocks, P. 1949. *Sickness in the Population of England and Wales in 1944-1947.* Studies on Medical and Population Subjects No. 2. Her Majesty's Stationery Office, London.

Stocks, P., and Campbell, J. M. 1955. Lung cancer death rates among nonsmokers and pipe and cigarette smokers. An evaluation in relation to air pollution by benzpyrene and other substances. *Brit. M. J., 2:* 923.

Strong, L. C. 1954. The cumulative effect of litter seriation on fibrosarcoma susceptibility in mice. *Ann. New York Acad. Sc., 57:* 507.

Sydenstricker, E. 1926. A study of illness in a general population group. *Pub. Health Rep.,* Sept. 24th.

Sydenstricker, E., Wheeler, G. A., and Goldberger, J. 1918. Disabling sickness among the population of seven cotton mill villages of South Carolina in relation to family income. *Pub. Health Rep., 33:* 47, Nov. 22nd.

Tetlow, C. 1955. Psychoses of childbearing. *J. Ment. Sc., 101:* 629.

Topley, W. W. C., and Wilson, G. S. 1936. *The Principles of Bacteriology and Immunity* (2nd ed.). William Wood & Co., Baltimore.

Trussell, R. E., Elinson, J., and Levin, M. L. 1956. Comparisons of various methods of estimating the prevalence of chronic disease in a community — the Hunterdon County study. *Am. J. Pub. Health, 46:* 173.

U. S. Bureau of the Census 1944, 1946, 1948. *Patients in Mental Institutions, 1941 (1943 & 1945), Part I. Patients in Hospitals for Mental Disease.* U. S. Govt. Printing Office, Washington, D. C.

U. S. Bureau of the Census 1949. *Historical Statistics of the United States, 1789-1945.* U. S. Govt. Printing Office, Washington, D. C.

U. S. Bureau of the Census 1953a. *United States Census of the Population: 1950. State of Birth.* Special Report P-E, No. 4A. U. S. Govt. Printing Office, Washington, D. C.

U. S. Bureau of Census 1953b. *United States Census of Population: 1950. Institutional Population.* Special Report P-E, No. 2c., U. S. Govt. Printing Office, Washington, D. C.

Van Buren, G. E. 1940. Some things you can't prove by mortality statistics. *Vital Statistics Special Reports, 12:* 3.

Wald, N. 1958. Leukemia in Hiroshima City atomic bomb survivors. *Science, 127:* 699.

Wheelis, J. M., Jr. 1938. A time-study of morbidity and mortality in the U. S. Navy. *Am. J. Pub. Health, 28:* 1291.

White, C. 1953. Sampling in medical research. *Brit. M. J., 2:* 1284.

White, C., and Bailar, J. C., III 1956. Retrospective and prospective methods of studying association in medicine. *Am. J. Pub. Health, 46:* 35.

Wiedorn, W. S., Jr. 1955. A unitary hypothesis of disorganisation: stress-organism relationships. *J. Nerv. & Ment. Dis., 122:* 575.

World Health Organization Monthly 1947. *Epidemiological and Vital Statistics Report, Vol. I* et seq., Geneva.

World Health Organization 1948. Text of the Constitution of the World Health Organization. *Off. Rec. World Health Organ., 2:* 100.

World Health Organization 1952. Comparability of statistics of causes of death according to the fifth and sixth revisions of the International List. *Bull. World Health Organ.,* Suppl. 4.

World Health Organization 1958. Poisoning in accidents, suicides and homicides; 1950-1956. *Epidemiol. & Vital Statist. Rep., 11:* 364.

Wynder, E. L. 1955. Environmental factors in cervical cancer: an approach to its prevention. *Brit. M. J., 1:* 743.

Wynder, E. L., and Lemon, F. R. 1958. Cancer, coronary artery disease and smoking. *California Med., 89:* 267.

Yerushalmy, J. 1938. Neonatal mortality by order of birth and age of parents. *Am. J. Hyg., 28:* 244.

Yerushalmy, J., and Hilleboe, H. E. 1957. Fat in the diet and mortality from heart disease. A methodologic note. *N. Y. State J. Med., 57:* 2343.

Zinsser, H. 1934. Varieties of typhus fever and the epidemiology of the American form of European typhus fever (Brill's disease). *Am. J. Hyg., 20:* 513.

Index

Absenteeism, measure of morbidity, 80
Accidents, automobile, 26
Accountants, seasonal variation in cholesterol, 128, 182
Accuracy of data, in case history studies, 248-252
Age. *See also* Age curves; Aging; Rates, age-specific
  as time specification, 65-68
  descriptive variable, 86-102
  diagnostic variation, 97
  parental. *See* Parental age
  purposes of study, 86
  rate of progression of disease, 97-98
Age curves. *See also* Rates, age-specific
  bimodality, 99
  cancer
    buccal cavity, 174-175
    carcinogenesis, 101
    lung, 94
    genetic selection, 100-101
    rectum, 87
  cohort, 88-95
    British mortality, 92
    interpretation, 95-97
  current, 88-95
  degenerative diseases, 99-102
  Hodgkin's disease, 99-100
  interpretation, 97-102
  multiple sclerosis, 98
  suicide, 122
  tuberculosis, 90, 91, 99
Age-specific rates. *See* Rates, age-specific
Aging, endogenous component, 101-102
Agreement, method of, 37
Air Pollution Medical Program, 151
Allison, A. C., 156
American Cancer Society, 213, 214, 215
American Public Health Association, 4
Analogy, method of, 37-38
Analytic epidemiology, 3, 32, 41-49, 210-267
  types of study, 44-48
Andvord, K. F., 88
Anencephaly
  familial occurrence, 27
  sex ratio, 27
Ankylosing spondylitis, leukemia in patients with, 217-218
Armed Forces
  deaths in, 75
  statistics of, 80
Arnold, F. A., Jr., 274
Association
  causal. *See also* Causation
    direct, 12, 16-17
    indirect, 12, 16-17
  secondary, 14-16
    in case history studies, 255-259
  statistical, 13-14
  types, 12
Ast, D. B., 274
Atomic Bomb Casualty Commission (U.S.), 212
Attack rate, 69
  secondary, 69
    seasonal variation, 181-182
Attacks, multiple, rates of, 70
Attributable case fatality rate, 57
Attributable risk
  definition, 229
  for community, 230
  in case history studies, 261
  in cohort studies, 229-230
Autopsy data, secular change, 168
Aycock, W. L., 183

Bailar, J. C., 45
Bailey, A. H., 132
Baker, G., 27

Bancroft, H., 255
Bantu, cancer, 108-109
BCG vaccination, 276-278
Beall, G., 101
Beeton, M., 205
Berman, C., 108
Bias, through vagaries of memory, 249-250
Bimodality, in age curves, 99
Biostatistics, relation to epidemiology, 10
Birth, as time specification, 65, 67
Birth order
  birth weight, 206-207
  comparison samples, 192-195
  comparisons with related population, 190-192
  definition, 189-190
  descriptive variable, 187-209
  family limitation, 203-204
  fertility, 203-204
  Greenwood-Yule method, 189, 195-201
  hospital records, 192
  interpretation, 204-208
  length of life, 205
  methods of determining association with, 188-203
  national statistics, 190-191
  neurosis, 195-198
  pyloric stenosis, 194-195, 201-202
  separation from parental age, 201-203
  social determinants of patterns, 206-207
  socio-economic status, 207
Birth rank. *See* Birth order
Birth weight
  birth order, 206-207
  litter size, 209
Broad Street pump, London cholera epidemic, 154-155, 164-165
Bronchitis, social class, 126

Cairns, M., 128
Campbell, J. M., 252
Cancer
  breast
    artificial menopause, 136
    fertility, 136
    marital status, 136
    secular change, 176-178
  buccal, secular changes, 173-175
  childhood, radiation and, 253-254
  liver, in Bantu, 108-109
  lung
    age curves, 93-95
    case history studies, 251, 254, 265
    cohort studies, 212, 213, 214, 215, 222, 224-225, 226, 230-232
Cancer (*continued*)
  lung (*continued*)
    epidemic, 4
    increase, 33, 37, 168
    mustard gas, 46
    nativity (New Zealand), 112-113
    Seventh-day Adventists, 41
    sex ratio, 33, 37
    smoking, 33, 37, 41, 226, 230-232, 251, 254, 265
  morbidity surveys
    Iowa, 82, 153
    ten U. S. metropolitan areas, 82, 87
    urban-rural comparisons, 153
  occupational, 127
    age at exposure, 101
  rectal, age and sex, 87-88
  religion and, 115
  scrotal, age at exposure, 101
  selected sites, secular change, 169
  stomach, in Japanese, 119
  thyroid, following thymus irradiation, 220-221
  uterine cervix
    fertility, 136-137
    Jewish women, 38, 115, 119-120
    marital status, 38, 136-137
    nuns, 36, 38
  uterine corpus
    fertility, 136-137
    marital status, 136-137
Cancer registers
  Connecticut, 78
  Denmark, 153
Caries, dental, 33, 37
  among inductees, 149, 150
  fluoride and, 22, 33, 37, 148, 274
Carter, C., 192
Case, R. A. M., 46, 92, 169, 218
Case fatality rate
  attributable, 57
  definition, 55
  related to incidence and death rate, 62, 74-75
Case history studies, 233-267
  accuracy of data, 248-252
  adjusted relative risk, 263-264
  advantages and disadvantages, 46-48
  analysis, 253-266
  attributable risk, 261
  bias, 47
  defined populations, 234, 239-241, 260-261
  definition, 46
  estimated risk, 259-266
  generalization from, 266-267

Case history studies (*continued*)
hospital and vital records, 248-249
interpretation, 266-267
interview bias, 251-252
matching and stratification, 244-247
memory, 249-250
nonresponse, 252-253
paired samples, 243-244
population comparison, 239-241
random and systematic sampling, 241-243
relative risk, 261-266
sampling frame for comparison groups, 236-239
secondary associations, 255-259
selection of cases, 234-235
selection of comparison groups, 235-247
sources of information, 247-252
statistical significance in, 255
Causation. *See also* Association, causal
additive, 19-20
attributable risk, 231
chains of, 18
concept, 11-12
consistency with existing knowledge, 15-16
definition, 12
demonstration, 15-16
experiment, 279
incidence, 62-63
necessary causes, 20
observational studies, 267
prevention and, 17, 21-22
relative risk, 231
special features, 12
statistical association as evidence, 13-14
strength of association, 15
therapy and, 22
time sequence, 15
web of, 18-21
Cause. *See* Causation; Association, causal
Caverly, C. S., 26
Census tracts, socio-economic classification, 129-130
Cerebral Palsy Collaborative Study, 214, 215, 223
Chapin, C. V., 132
Cholera, 33, 37
epidemic in Broad Street, 154-155, 164-165
London water supply, 40, 42-44
Cholesterol, serum
accountants, 128, 182
Neapolitan and American males, 144
Ciocco, A., 139
Classification
classified, 23-24
defined, 23
Linnaeus' system, 23
purposes, 23
Classification of ill persons, 23-31
criteria for, 24
lack of correspondence, 25-26
descriptive epidemiology and, 26-27, 99
experiential, 24
some postulates for, 30-31
manifestational, 24
microorganisms and, 27-30
Cleft palate, sex ratio, 27
Clemmesen, J., 153
Clinical medicine, relation to epidemiology, 10
Clustering in time, 182-186
Cochran, W. G., 239
Cohort analysis
of age curves, 88-95
of secular change, 173-179
Cohort patterns, interpretation, 95-97, 173-179
Cohort studies, 210-232
advantages and disadvantages, 46-48
analysis, 223-230
clinical examination in, 223
comparison groups, 218-221
comparison with general population, 219
data on exposure, 214-217
death as outcome, 222
definition, 46
follow-up, 221-223
general populations
difficulties, 211
types, 211-217
insured persons, 213
interpretation, 230-232
loss to follow-up, 222, 227-229
mail questionnaires, 215
medical care groups, 213
obstetric populations, 214
occupational, 212-213
physical examination, 216-217
relative and attributable risk, 229-230
self-selected groups, 213-214
smoking, 212, 213, 214, 215, 222, 225, 226, 230-232
special exposure groups, 217-221
types, 210
whole communities, 212
Collins, S. D., 83, 107, 111
Commission on Chronic Illness, 83
Communicable Disease Center, 78

Comparison groups
for birth order and parental age, 192-195
in case history studies
associates, 239
general population, 239-241
matching, 244-247
other patients, 236-238
relatives, 238-239
sampling frame, 236-239
stratification, 244-247
in cohort studies, 218-221
neighborhood, 243-244
paired samples in case history studies, 243-244
random and systematic sampling, 241-243
sampling and, 239-247
Concomitant variation, method of, 37
Connecticut Cancer Register, 78
Controls. *See* Comparison groups
Copenhagen, tuberculosis register, 78-80
Coronary heart disease
conditions of work, 128
diet, 143-144
epidemic, 4
Framingham study, 212, 216, 222
social class, 126
Countries, variations in disease patterns, 144-153
Court-Brown, W. M., 217, 218, 219
Croxton, F. E., 255
Cullen, W., 29
Cutler, S. J., 82, 87, 94
Cyclic fluctuations in disease rates, 179-182

Danish Cancer Registry, urban-rural comparisons, 153
Dawber, T. R., 212
Day, A., 132
Dean, H. T., 148
Death. *See also* Death rate; Death Registration Area
cause
International List, 73-74
judgment as to, 56-57
multiple, 74
statistical assessment, 56-57
underlying, 74
registration, 72-74
changes in, 167, 176
season, 179-181
Death rate
definition, 55
disparity between sexes, 104, 105
fetal. *See* Fetal death rate
Death rate (*continued*)
index of incidence, 57, 63
leukemia (example), 60
marital status, 134-135
prediction, 178-179
race-specific, various causes, 111
related to incidence and case fatality, 62, 74-75
sex-specific, various causes, 105, 106
social class, 126
Death ratio, fetal, 53
Death Registration Area, 73
Decreasing incidence, effect on age curves, 95-96
Definitions
analytic epidemiology, 3, 32
attributable risk, 229
birth order, 189-190
case fatality rate, 55
case history study, 46, 233
causation, 12
classification, 23
cohort study, 46, 210
death rate, 55
descriptive epidemiology, 3, 32
epidemic, 4
epidemiology, 3
equivalent average death rate, 137
ethnic group, 107-108
experimental epidemiology, 270
fetal death rate, 53
fetal death ratio, 53
incidence, 54-55
period prevalence, 54
point prevalence, 54
rate, 52
ratio, 53
relative risk, 229
sex ratio, 103
standardized mortality ratio, 126
Degenerative diseases, age curves, 99-102
Descriptive epidemiology, 3, 32, 34-35
Descriptive variables
age, 86-102
birth order, 187-209
ethnic group, 107-123
fertility, 140
indirect associations with, 118-119
interpretation of disease patterns, 116-123
litter size, 208-209
marital status, 134-140
nativity, 112-113
occupation, 124-128
parental age, 187-209
place, 141-163

Descriptive variables (*continued*)
race, 109-112
religion, 113-116
season, 179-182
sex, 103-107
socio-economic status, 128-134
time, 164-186
Devonshire colic, 27
Diabetes mellitus, 22, 74
Diagnosis, and secular change, 167-170, 176
Diet, and coronary heart disease, 143-144
Difference, method of, 36
Diphtheria
immunization, 269
secular trend, 269
Disability. *See* Disease
Disabled persons. *See* Classification of ill persons
Disease
delimitation of entities
by experiential criteria, 24-26, 27-31
by manifestational criteria, 24-27
entity, 31
frequency. *See also* Incidence; Prevalence; Death rate
administrative changes, 79-80
inception, 21-22
registers, 78-80
reportable, international, 77
reporting, 77-78
Doll, R., 48, 212, 215, 217, 218, 219, 222, 224, 225, 226, 230, 231, 246, 251, 254, 256
Dorn, H. F., 72, 82, 87, 94, 213, 215, 222, 224, 227
Double-blind method, 272-273
Downes, J., 83, 107, 111, 139
Doyle, J. T., 212, 216
Dublin, L. I., 205
Duncan, J. M., 207
Dunn, H. L., 74
Dunning, J. M., 150
Duration of illness
leukemia (example), 61
relation to incidence and prevalence, 60-62

Eastcott, D. F., 113
Eaton, J. W., 116
Economic status. *See* Socio-economic status
Economic subregions of U. S., 144, 147
Eldering, G., 271, 272
Enamel, mottled, 33
Enos, W. F., Jr., 57
Enteric disease, 17, 22
Entity, disease, 31
Environment
biological, peculiar to place, 158
chemical and physical, peculiar to place, 158-159
cultural
ethnic group differences, 120
peculiar to place, 159
intrauterine, 187-188
litter size and, 208-209
sex differences in, 120
Epidemic
cancer of lung, 4
concept, 4-7
coronary heart disease, 4
definition, 4
"point," 164-166
recognition, 4-5
reporting, 77-78
Epidemiology
administrative purposes, 9
analytic, 3, 32, 41-49, 210-267
anecdotal, 141
definition, 3
descriptive, 3, 32, 34-35
experimental, 32, 49-50, 268-279
"practical," 8
purposes, 7-9
relation to clinical medicine, 10
relation to disease prevention, 7
relation to other disciplines, 9-10
strategy, 32-50
use in classification of ill persons, 7, 23-31
Equivalent average death rate, 137
Erhardt, C. L., 74
Ethnic group
definition, 107-108
descriptive variable, 107-123
geographic patterns, 119
interpretation, 116-123
place disease, 143-144
types, 108
Events
categories, 12
specification in rates, 69-70
Experiential criteria, in delimiting disease entities, 24, 27-31
Experiment. *See also* Experimental epidemiology
controlled, 270-276
"natural," 276-278
uncontrolled, 49-50, 268-270
Experimental epidemiology, 32, 49-50, 268-279
definition, 270
double-blind method, 272-273

Experimental epidemiology (*continued*)
ethics, 274
placebos, 272
practice and, 268-270, 275-276, 279
randomization in, 272-274
selection of population, 270-271
sequential designs, 274-275
study design, 271-274

FAMILIAL DISEASE, 38
Greenwood-Yule method, 199
Family limitation, effect on birth order associations, 203-204
Family size, effect on birth order associations, 203
Feinleib, M., 136
Fertility
descriptive variable, 140
effect on birth order and parental age associations, 203-204
Fetal death rate
definition, 53
race and birth order, 191
Fetal death ratio, 53
Fluoride in water
dental caries, 22, 33, 37, 148, 274
experiments, 274
Fog, deaths in London, 5, 6, 75
Fogh-Andersen, P., 27
Follow-up, in cohort studies, 221-223
Follow-up loss
in cohort studies, 222, 227-229
treatment of, 228-229
Food rationing, 81
Forlano, G., 182
Framingham heart study, 212, 216, 222
Francis, T., Jr., 276
Fraser, R., 128
Friedman, M., 128, 182
Frost, W. H., 88, 89, 96

GALTON, F., 207
Gamma-globulin, and poliomyelitis, 275-276
Garratt, F. N., 130, 131
Gastric ulcer, social class, 126
General paralysis and syphilis, geographic distribution, 158
Genetic constitution, related to epidemiologic patterns, 121-122
Gentry, J. T., 149, 159
Geographic variation. *See also* Place; Place associations; Place diseases
leukemia in U.S., 144-147
Gilliam, A. G., 108, 145, 147, 170
Goldberger, J., 133
Gordon, J. E., 89
Greenwood, M., 49, 189
Greenwood-Yule method
birth order, 189, 195-201
deceased siblings, 199
familial disorders, 199
incomplete sibships, 199-201
parental age, 189, 201
twins, 198-199
Gregg, N. M., 187
Griswold, M. H., 78

HADLEY, J. N., 116
Haenszel, W., 82, 153, 255, 259, 264
Hammon, W. McD., 275
Hammond, E. C., 213
Harelip, sex ratio, 27
Health
positive, 8
W.H.O. definition, 8
Health Insurance Plan of Greater New York, 213
Heart disease, coronary. *See* Coronary heart disease
Hempel, C. G., 23
Hepatitis
serum, 14-15, 16-17, 19
yellow fever vaccination, 184-185
Hill, A. B., 47, 48, 212, 215, 222, 224, 225, 226, 230, 231, 246, 251, 254, 255, 256
Hilleboe, H. E., 143
Hirsch, A., 141, 159
Hodgkin's disease
age curve, 99-100
classification, 99
Hollingshead, A. B., 132
Holt, S. B., 204
Horn, D., 213
Hospital records, birth order and parental age, 192
Hueper, W. C., 101, 102, 127
Hume, D., 11
Hunterdon County morbidity survey, 83
sources of data, 216
Hutterites, 116
Huxley, J., 100
Hyge, T. V., 276, 278
Hypertension, conditions of work, 128
Hypotheses
evaluation, 38-41
formulation, 32, 36-38

ILL PERSONS. *See* Classification of ill persons
Inaccuracy of data, implications, 248
Inception, disease, 21-22
Incidence
death rates as index, 57, 63

Incidence (*continued*)
definition, 54-55
leukemia (example), 59-60
population at risk, 70-71
relation to death and case fatality rates, 62, 74-75
relation to point prevalence and duration, 60-62
total, 69
uses of, 62-63
Income, as measure of socio-economic status, 129
Increasing incidence, effect on age curves, 95-96
Indian, American, 116
Inductees
dental caries, 149, 150
sarcoidosis, 149
Infant mortality
overcrowding, 131-132
social class, 127
Infectivity, seasonal variation, 181
Influenza
epidemic of *1918-1919*, 92
epidemic of *1957*, 75-76, 78, 84
Inheritance, nongenic, 205
Insurance, death data, 75
Intelligence, and season of birth, 182
International List, 73-74
International variation in disease rates, 142-144
Interview bias, in case history studies, 251-252
International Sanitary Regulation, 77
Ipsen, J., Jr., 82, 98
Irradiation. *See* Radiation, ionizing

JALAVISTO, E., 205

KENDRICK, P., 271, 272
Kettle, E. H., 30
Keys, A., 144
Knobloch, H., 182
Koch, R., 27, 28, 29
postulates, 27-28
Korteweg, R., 93
Kraepelin, E., 158
Kutschenreuter, P. H., 180

LANDAUER, W., 101
Lange, R. D., 212
Lansing, A. I., 205
Lea, A. J., 46, 218
Lee, E. S., 162, 163
Lee, R. E., 128
Leighton, A., 83
Lemon, F. R., 41
Leukemia
case fatality rate, 62
death rate, 60
duration of illness, 61
geographic variation in U.S., 144-147
incidence, 59
prevalence, 58
radiation
ankylosing spondylitis, 217-218
atomic bombs, 48, 212
thymic enlargement, 220-221
social class, 126
Life time probability, 69
Lilienfeld, A. M., 136
Linder, F. E., 84
Litter size. *See also* Twins
descriptive variable, 208-209
Livingstone, F. B., 157
Logan, W. P. D., 137
London fog, deaths, 5, 6, 75
Lowe, C. R., 38, 130, 131, 257, 258, 264
Luther, E. H., 183

MACCARTHY, D., 192
MacMahon, B., 27, 68, 99, 100, 136, 174, 176
Mainland, D., 242
Malformations, central nervous system, 252-253
Malzberg, B., 113, 162, 163
Manic-depressive disease, recency of migration, 162
Manifestational criteria, in delimiting disease entities, 24-27
Manos, N. E., 151, 152
Mantel, N., 255, 259, 264
Manual of Joint Causes of Death, 74
Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death, 73-74
Marital partners, concordance of disease, 139-140
Marital status
descriptive variable, 134-140
interpretation of disease patterns, 135
mental disease, 136-139
mortality rates, 134-135
Marks, H. H., 205
Martin, W. J., 131, 132
Masland, R. L., 214
Massachusetts
diphtheria, 268-269
pertussis, 268-269
Matching, in case history studies, 244-247
Maternal age. *See* Parental age

Maternal deprivation, mental illness, 48
McDonald, A. D., 214
McKeown, T., 193, 194, 200, 202, 206, 245, 246, 252
Measurement
  disease frequency, 51-71
  errors of, 117-118
Medical Research Council (British), 271, 275
Medicine, clinical, relation to epidemiology, 10
Memory
  as source of data, 249-250
  checks of, 250-252
Menaker, W., 263
Menopause, artificial, and cancer of breast, 136
Mental disease. *See also* Mental hospital admissions
  Faroe Islands, 160
  neurosis and birth order, 195-198
  pregnancy and puerperium, 183, 184
Mental hospital admissions
  economic status, 130-131
  marital status, 137-139
  military stress, 171-173
  nativity, 113
  secular change, 171-173
  socio-economic status, 133
Mental hospital statistics, 81
Mental retardation, season of birth, 182
Methodism, Hirsch on, 159
Microbiology, relation to epidemiology, 10
Microorganism
  classification around, 27-30
  internalization, 27-30
Migration
  effects, 160-163
  infectious disease, 160
  nativity, 160
  psychosis, 160-163
  racial, within U.S., 161
Mill, J. S., 31
Miller, D. C., 144
Ministry of Health, 6
Mongolism, maternal age, 192
Morbidity
  absenteeism as index, 80
  income, 130
  reporting, 77-81
  sex, 105-107
  sources of data, 72-85
  surveys, 81-85, 107
    British, 83
    National Health Survey (U.S.), 84-85, 106, 107, 111, 118, 129, 130
Morbidity (*continued*)
    other U.S., 83
Morris, J. N., 128
Mortality. *See also* Death rate; Infant mortality; Standardized mortality ratio
  sources of data, 72-85
Mother death, during infancy, 30
Muench, H., 57
Multiple sclerosis
  age curves, 98
  morbidity survey in Boston, 82, 98
Mustard gas poisoning, cohort study, 46, 218

National Cancer Institute, 246
National Health Survey, 84-85, 106, 107, 111, 118, 129, 130
National Institute of Neurological Diseases and Blindness, 214
National Office of Vital Statistics, 74, 75, 76, 134, 135, 149, 191
Nativity
  descriptive variable, 112-113
  measure of migration, 160
"Natural" experiments, 276-278
Neoplasm. *See* Cancer
Neurosis
  birth order, 195-198
  conditions of work, 128
Nielsen, A., 153
Nonresponse, in case history studies, 252-253
Norton, A., 195, 196, 197
Novitski, E., 202

Observation, epidemiologic, 10, 279
Occupation
  descriptive variable, 124-128
  epidemiologic uses, 124-125
  Registrar General's classification, 126
  socio-economic status, 125-127, 129
  special risks, 127
  work conditions, 128
Opler, M. K., 83
Outcome, measures in cohort studies, 221-223

Panum, P. L., 41, 159
Parasitic diseases, as place diseases, 143
Parental age
  comparison groups, 192-195
  comparison with population, 190-192
  descriptive variable, 187-209
  Greenwood-Yule method, 201. *See also* Greenwood-Yule method

Parental age (*continued*)
hospital records, 192
interpretation, 204-208
methods of determining association with, 188-203
mongolism, 192
national statistics, 190
pyloric stenosis, 201-202
separation from birth order, 201-203
social determinants of patterns, 206-207
Parity, maternal. *See* Birth order
Pasamanick, B., 182
Paternal age. *See* Parental age
Pathology, relation to epidemiology, 10
Pearson, J. T., 169
Pearson, K., 205
Pellagra, socio-economic status, 133
Penrose, L. S., 203
Period prevalence
definition, 54
leukemia (example), 59, 60
uses of, 64
Perrott, G. St. J., 83
Persons, epidemiologic characteristics, 35, 86-140
Person-years of risk
calculation, 224-225
difficulties, 225-227
Pertussis
vaccination, 269, 271-272, 275
secular trend, 269
Pintner, R., 182
Place. *See also* Place diseases
changes in. *See* Migration
characteristics, 35
descriptive variable, 141-163
international comparisons, 142-144
local disease patterns, 154-156
methods of measurement, 141-156
Place diseases. *See also* Place
characteristics, 157
due to environment, 157-160
due to inhabitants, 156-157
ethnic group, 143-144
explanation, 156-160
Place in family. *See* Birth order
Placebo, use in experiments, 272
Point epidemics, 164-166
Point prevalence
at birth, 67
at death, 57
definition, 54
estimation from death rate, 74-75
inductees, 68
leukemia (example), 58-59, 60
Point prevalence (*continued*)
related or reference population, 52-53
relation to incidence and duration, 60-62
uses of, 63
Poliomyelitis
gamma-globulin, 275-276
racial differences, 121
tonsillectomy, 183
vaccine trials, 276
Pollard, A. H., 178
Pope, A. S., 89
Population
at risk, 52-53, 70-71
related or reference, 52-53, 70-71
specification in rates, 70-71
Postulates, Koch's, 27-28
Prediction
death rates, 178-179
time trends, 176-179
Pregnancy
cohort studies, 48
mental disease, 183-184
Prevalence. *See* Period prevalence; Point prevalence
Prevention, side effects, 22
Proportional mortality, error of interpretation, 108-109
Prospective studies, 45-46
Psychosis. *See* Mental disease; Mental hospital admissions
Puerperium, mental disease in, 183-184
Pyloric stenosis of infancy
birth order, 194-195, 201-202
incidence, 68
parental age, 201-202

Race. *See also* Ethnic group
descriptive variable, 109-112
mortality rates, 109-112
Radiation, ionizing
childhood cancer, 253-254, 260-261
leukemia. *See* Leukemia
Rate. *See also* Case fatality rate; Death rate; Incidence; Period prevalence; Point prevalence
definition, 52
disease frequency, 51-71
fetal death, 53, 191
Rates
age-specific, 86-88. *See also* Age curves
cancer
breast, 177
buccal cavity, 174-175
lung, 94
rectum, 87

Rates (*continued*)
age-specific (*continued*)
Hodgkin's disease, 100
mortality, 104
by race, 109
multiple sclerosis, 98
suicide, 122
tuberculosis, 89-91
age-standardized, 118-119
mortality, by race, 109-112
sex-specific
cancer
breast, 177
lung, 94
rectum, 87
morbidity, 107
mortality, 104-105
by cause, 106
suicide, 122
tuberculosis, 89-91
Ratio. *See also* Standardized mortality ratio
definition, 53
fetal death, 53
Record, R. G., 193, 200, 245, 246, 252
Recovery rate, 55-56
Redlich, F. C., 132
Reed, S. C., 204
Registrar General (England and Wales), 74, 84, 125, 149, 151, 177
occupational classes, 126
Occupational Supplements, 126, 127
Relative risk
adjusted, 263-264
case history studies, 261-266
cohort studies, 229-230
definition, 229
index of causal association, 231
multiple groupings of study variables, 264-265
Relatives, as comparison group, 238-239
Religion
descriptive variable, 113-116
interpretation of disease associations with, 114
relation to ethnic group, 113-114
Retrospective studies, 45-46
Residence area, as measure of socioeconomic status, 129-132
Risk. *See also* Attributable risk; Relative risk
estimation in case history studies, 259-266
person-years of, 224-227
population at, 52-53, 70-71
Sampling. *See also* Comparison groups
case history studies, 239-247
from lists in which individuals appear more than once, 242-243
paired, 243-244
random and systematic, 241-243
Sampling frame, choice in case history studies, 236-239
Sandler, L., 202
Sarcoidosis
geographic localization, 149
inductees, 149
Sartwell, P. E., 219
Sawyer, W. A., 184, 185
Schizophrenia, and recency of migration, 162. *See also* Mental disease; Mental hospital admissions
Schneider, R. F., 128
Scrimshaw, N. S., 143
Season
descriptive variable, 179-182
interpretation of disease patterns, 179-182
of birth, and intelligence, 182
of death, 76, 179-181
Secondary attack rate, 69
Secular change, 166-176
artificial causes, 167-171
assessment of role of complicating factors, 168-171, 176
autopsy data, 168
cancer
breast, 176-178
buccal cavity, 173-175
lung, 168-170
selected sites, 169
cohort analysis, 173-176
death registration procedures, 167, 176
diagnosis, 167-170, 176
diphtheria, 268-269
interpretation, 171-173, 176
mental hospital admissions, 171-173
pertussis, 268-269
population characteristics, 167
prediction, 176-179
rapid, 40-41, 171-173
Seltser, R., 219
Sequential design of experiments, 274-275
Serum hepatitis, 14-15, 16-17, 19
yellow fever vaccination, 184-185
Seventh-day Adventists, lung cancer, 41
Sex
descriptive variable, 103-107
interpretation of disease differences, 116-123

Sex linkage, 121
Sex-specific rates. *See* Rates, sex-specific
Sex ratio
  definition, 103
  at birth, 104
  use to measure sex association, 103-104
Shackley, W., 74
Shope, R. E., 30
Siblings, as comparison group, 238
Sickle-cell trait, and malaria, 156-157
Side effects of prevention, 22
Silica suspension, intratracheal, 30
Silicotuberculosis, 30
Simpson, C. L., 217, 219, 220, 221
Simpson, R. E. H., 181
Slater, P., 83
Smith, R. L., 112, 116, 119
Smoking
  case history studies, 251, 254, 257-258, 265-266
  cohort studies, 212, 213, 214, 215, 222, 225, 226, 230-232
  deaths from selected causes, 231
  lung cancer, 33, 37, 41, 226, 230-232, 251, 254, 265
  tuberculosis, 38, 257-258, 264, 265-266
Snow, J., 5, 33, 40, 41, 42-44, 154, 155, 165, 171
Social class. *See* Socio-economic status
Social stress, internalization of, 27
Socio-economic status
  area of residence, 129-132
  complex measures, 132
  descriptive variable, 128-134
  income as measure, 129
  interpretation of disease patterns, 132-134
  measurement, 128-132
  occupation as measure, 125-127, 129
  reproductive patterns, 207
Spina bifida
  familial occurrence, 27
  sex ratio, 27
Spot map, 154
Spouse
  comparison in case history studies, 238
  same disease in, 139-140
Springett, V. H., 93
Standard Metropolitan Areas of U.S., 151
  death rates, 151-153
Standardized mortality ratio
  definition, 126
  geographic areas, 144-147
  social class, 126
Statistical significance, case history studies, 255
Stewart, A., 128, 254
Still, G. F., 194, 195
Stocks, P., 81, 84, 107, 252
Stratification, in case history studies, 244-247
Stress
  military, admission to mental hospital, 171-173
  social, internalization of, 27
Strong, L. C., 205
Suicide, Japan and U.S.A., 122
Surveys of morbidity, 81-85, 107
Swine influenza, 30
Sydenstricker, E., 83
Syphilis
  and general paralysis, geographic distribution, 158
  icterus and, 14-15, 16-17, 19

TEMPERATURE, monthly deaths, 179-181
Terry, W. D., 174, 176
Tetlow, C., 183, 184
Thymus, irradiation, 220-221
Thyroid. *See* Cancer, thyroid
Time
  clustering in, 182-186
  descriptive variable, 164-186
  points in, 65-68
  rapid changes, 40
  specification in rates, 64-69
  specification of lengths, 68-69
  variables describing, 35
Time trends, prediction, 176-179
Tobacco. *See* Smoking
Tonsillectomy, and poliomyelitis, 183
Topley, W. W. C., 27, 28, 49
Triplets, birth weight, 209
Trussell, R. E., 216
Tuberculosis
  bimodal age curve, 99
  classification, 20, 28-29
  cohort age curves, 91
  conditions of work, 128
  current age curves, 90
  implications of cohort pattern, 96-97
  prevalence estimation, 75
  register, Copenhagen, 78-80
  school epidemic, 276-278
  smoking, 38, 257-258, 264, 265-266
  social class, 126
Typhus, in Jewish population of New York City, 112, 114
Twins
  birth weight, 209
  expected frequency in affected, 208-209
  Greenwood-Yule method, 198-199

URBAN-RURAL CLASSIFICATION
England and Wales, 151
U.S.A., 150-151
Urban-rural comparison
cancer, 153
death rates, 151-153
disease frequency, 149-153
infectious diseases, 151-152
U.S. Bureau of the Census, 138, 161, 172
U.S. Public Health Service, 82, 83, 151, 274. *See also* National Cancer Institute; National Health Survey; National Institute of Neurological Diseases and Blindness; National Office of Vital Statistics

VACCINATION, yellow fever and hepatitis, 184-185
Van Buren, G. E., 74
Variation, method of concomitant, 37
Vitamin C, 17

WALD, N., 48
Water supply, cholera, 40, 42-44, 154-155
Web of causation, 18-21
Weil, R. J., 116
Weiner, L., 74
Wheelis, J. M., Jr., 160
White, C., 45, 242
Whooping cough. *See* Pertussis
Wiedorn, W. S., Jr., 30
Wilson, G. S., 27, 28, 49
World Health Organization, 8, 74, 122, 142
Wynder, E. L., 41, 140

YELLOW FEVER, vaccination, hepatitis and, 184-185
Yerushalmy, J., 143, 202
Yule, G. U., 189

ZINSSER, H., 112, 114

Neighborhood map / relative - good for matching but
not representative of general population because Can't use
for incidence estimation

DEPT. OF COMMUNITY MEDICINE
BANGALORE-34
ST. JOHN'S MEDICAL COLLEGE

Community Health Cell
Library and Documentation Unit
367, "Srinivasa Nilaya"
Jakkasandra 1st Main,
1st Block, Koramangala,
BANGALORE-560 034.
Phone : 5531518